REMARKS

Claims 1 through 19 are pending in the application of which claims 3 through 6 have been withdrawn from consideration. Claim 2, and corresponding dependent claims 10 through 19, have been rejected under 35 U.S.C. §101. Claim 1, and corresponding dependent claims 2, and 7 through 19, have been rejected under 35 U.S.C. §112.

Amendments to the specification

The paragraph beginning on the last two lines of page 34 and continuing on page 35 has been amended to remove a reference to a website.

Amendments to the claims

Claim 1 has been amended to clarify the invention. Support can be found in the specification, inter alia, at page 8, lines 13-18, and in the claims as originally filed. Claim 2 has been amended to clarify the invention. Support can be found, inter alia, in the claims as originally filed.

Claims rejected under 35 U.S.C. §101

The Examiner has stated that "[c]laims 2 and 10-19 are drawn to a host cell, which reads on a product of nature." The Applicant has previously argued that claim 2 does not describe a product of nature, because the host cell comprises recombinant genes, which are not products found in nature. As defined in the specification (p. 6, lines 2-3), "...the term recombinant refers to a cell, compound, or composition produced at least in part by human intervention."

Nonetheless, the Examiner remains concerned that claim 2 reads on a product of nature. Though Applicant does not necessarily agree with the Examiner, Applicant has amended claim 2 to address her concerns by inserting the phrase "an isolated" in front of "host cell". Accordingly, Applicant respectfully requests that the Examiner withdraw the 35 U.S.C. §101 rejection of amended claim 2, and the corresponding dependent claims 10 through 19.

<u>Claims rejected under 35 U.S.C. §112</u>

Claim 1, and corresponding dependent claims 2, and 7 through 19, have been rejected under 35 U.S.C. §112 as "containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention."

In response, Applicant previously clarified that the subject matter at issue were known structures at the time of the invention. Therefore, where the prior art teaches what the applicant refers to, no sequence or structural description is needed in the specification, let alone in the claims. Neither recitation of the sequences nor incorporation by reference is necessary where what the application refers to is known.

To further address the Examiner's concerns, Applicant offers the following references to illustrate that the biological sequences used in the claimed invention were known individually in the art at the time of the invention:

Claim Element	GenBank Reference
KS^Q domain	GenBank AF016585 (12/7/1997)
AT domain	GenBank AF016585 (12/7/1997)
ACP domain	GenBank X56107 (10/14/1991)
ccr	GenBank accession no. U67612 (3/31/1998)
icm	GenBank accession no. U37135 (3/27/1996)

Additionally, scientific literature references include:

KS^Q domain

> Kakavas, S. J., *et al.*, "Identification and Characterization of the Niddamycin Polyketide Synthase Genes from *Streptomyces caelestis*," <u>Journal of Bacteriology</u>, Dec. 1997, pp. 7515-7522. (Exhibit A)

AT domain

>   Stassi, D. L., *et al.*, "Ethyl-substituted erythromycin derivatives produced by
>   directed metabolic engineering," <u>Proc. Natl. Acad. Sci. USA</u>, Vol. 95, pp.
>   7305-7309, June 1998. (Exhibit B)

ACP domain

>   Donadio, S. and Katz, L., "Organization of the enzymatic domains in the
>   multifunctional polyketide synthase involved in erythromycine formation in
>   *Saccharopolyspora erythraea*," <u>Gene</u>, 111 (1992) pp. 51-60. (Exhibit C)

ccr

>   Vrijbloed, J.W., *et al.*, "Insertional Inactivation of Methylmalonyl Coenzyme
>   A (CoA) Mutase and Isobutyryl-CoA Mutase Genes in *Streptomyces
>   cinnamonensis*: Influence of Polyketide Antibiotic Biosynthesis," <u>Journal of
>   Bacteriology</u>, Sept. 1999 p. 5600-5605. (Exhibit D)

icm

>   Stassi, D. L., *et al.*, "Ethyl-substituted erythromycin derivatives produced by
>   directed metabolic engineering," <u>Proc. Natl. Acad. Sci. USA</u>, Vol. 95, pp.
>   7305-7309, June 1998. (Exhibit E)

   Applicant respectfully requests that the Examiner considers the above references as
supporting the Applicant's assertion that the individual claim elements of claims 1 and 2, and their
corresponding dependent claims, were well known at the time that the application was filed.
Accordingly, the Examiner is respectfully requested to withdraw the 35 U.S.C. §112 rejection.

CONCLUSION

In view of the above, each of the presently pending claims in this application is believed to be in immediate condition for allowance. Accordingly, the Examiner is respectfully requested to withdraw the outstanding rejection of the claims and to pass this application to issue.

If it is determined that a telephone conference would expedite the prosecution of this application, the Examiner is invited to telephone the undersigned at the number given below.

In the event the U.S. Patent and Trademark office determines that an extension and/or other relief is required, applicant petitions for any required relief including extensions of time and authorizes the Commissioner to charge the cost of such petitions and/or other fees due in connection with the filing of this document to Deposit Account No. 03-1952 referencing docket no. 30062-2004810. However, the Commissioner is not authorized to charge the cost of the issue fee to the Deposit Account.

Dated: February 16, 2007   Respectfully submitted,

By _Brenda J Wallach_
Brenda J. Wallach, Ph.D.
   Registration No.: 45,193
MORRISON & FOERSTER LLP
12531 High Bluff Drive, Suite 100
San Diego, California 92130-2040
(858)720-7961

sd-357877

Identification and Characterization of the Niddamycin Polyketide Synthase Genes from *Streptomyces caelestis*

STEPHAN J. KAKAVAS, LEONARD KATZ, AND DIANE STASSI*

Antibacterial Discovery Division, Abbott Laboratories, Abbott Park, Illinois 60064

The genes encoding the polyketide synthase (PKS) portion of the niddamycin biosynthetic pathway were isolated from a library of *Streptomyces caelestis* NRRL-2821 chromosomal DNA. Analysis of 40 kb of DNA revealed the presence of five large open reading frames (ORFs) encoding the seven modular sets of enzymatic activities required for the synthesis of a 16-membered lactone ring. The enzymatic motifs identified within each module were consistent with those predicted from the structure of niddamycin. Disruption of the second ORF of the PKS coding region eliminated niddamycin production, demonstrating that the cloned genes are involved in the biosynthesis of this compound.

Niddamycin is a macrolide antibiotic which is able to bind 50S ribosomal subunits to inhibit protein synthesis. The compound was first discovered as a secondary metabolite of *Streptomyces djakartensis* (16) and was later found to be produced by *Streptomyces caelestis* NRRL-2821 (11a). The structure of niddamycin (Fig. 1) suggests that the polyketide backbone of the macrolide ring is formed through the ordered condensation of carboxylic acid residues derived from acetate, propionate, butyrate, and perhaps glycolate (24). The disaccharide, mycaminose-isobutyrylmycarose, is attached to the macrolide ring at C-5.

Macrolides belong to a class of molecules referred to as complex polyketides, which are synthesized on large, multifunctional enzymes called polyketide synthases (PKSs). The synthesis of polyketides is mechanistically similar to that of fatty acids; however, a greater variety of starter and extender carboxylic acid residues are incorporated into the growing polyketide chain, and the β-keto groups formed after each condensation step undergo various degrees of reduction (15, 20).

PKSs, in general, contain all of the enzymatic activities necessary for the sequential condensation of acyl thioesters (β-ketoacyl acyl carrier protein synthases [KS], acyltransferases [AT], and acyl carrier proteins [ACP]), the subsequent reduction of the β-keto groups (dehydratases [DH], enoylreductases [ER], and ketoreductases [KR]), and the release of the completed chain from the PKS (thioesterases [TE]). Analysis of the erythromycin PKS genes revealed that these enzymatic domains are organized into modules, each of which is responsible for one round of condensation and reduction (5, 7, 9, 10). As a result, there is a direct correlation between the number of modules contained within the erythromycin PKS and the length of the polyketide chain. In addition, the genetic order of the erythromycin PKS modules was found to be colinear with the order of biochemical reactions, allowing directed genetic alterations which produce predicted novel erythromycin derivatives (9, 11).

The polyketide portion of the 16-membered macrolide niddamycin is predicted to be synthesized by a complex (type 1) PKS (15) comprising seven modules, each catalyzing one con-

densation reaction. It had previously been suggested that the choice of the extender coenzyme A (CoA)-thioester is determined by the AT domain contained in each module (9). Sequence comparisons of AT domains for fatty acid and polyketide synthases have revealed specific sequence motifs for malonyl- and methylmalonyl-ATs (m- and mmATs) (13). The chemical structure of niddamycin (Fig. 1) suggests that the PKS should contain ATs specific not only for malonyl-CoA (modules 1, 2, 3, and 7) and methylmalonyl-CoA (module 4) but also for the more rarely encountered ethylmalonyl-CoA (module 5) and for an as yet undetermined CoA derivative which results in the insertion of hydroxymalonate into the growing chain (module 6). In addition, the structure predicts the enzymatic motifs which should be present in each module to give the corresponding reduction state of the β-carbonyl group formed after each condensation.

This study describes the isolation and characterization of the PKS genes responsible for niddamycin biosynthesis, including those AT sequences which may result in the introduction of unusual side chains to the macrolide ring.

MATERIALS AND METHODS

Media and reagents. SeaKem-agarose was obtained from FMC BioProducts, Rockland, Maine. Bacto Soytone, soluble starch, yeast extract, Bacto Agar, and Antibiotic Medium 11 (AM11) were obtained from Difco Laboratories, Detroit, Mich. SGGP medium (34) and AS-1 medium (4) have been described previously. SCM medium contains (per liter) the following: Bacto Soytone, 20 g; soluble starch, 15 g; morpholinepropanesulfonic acid 10.5 g; yeast extract, 1.5 g; and CaCl₂, 0.1 g. R3M plates consist of (per liter) the following: sucrose, 103 g; K₂SO₄, 0.25 g; yeast extract, 4 g; Casamino Acids, 4 g; tryptone, 4 g; agar, 22 g; and H₂O, 830 ml. After sterilization, 20 ml of a 2.5 M solution of MgCl₂, 20 ml of a 50% glucose solution, 20 ml of a 2.5 M solution of CaCl₂, 12.5 ml of a 2 M solution of Tris-HCl (pH 7.0), 0.2 ml of a 5,000× concentrated trace element solution (14), 2.5 ml of a 1 M solution of NaOH, and 0.37 ml of a 0.5 M solution of KH₂PO₄ were added.

Strains, bacteriophage, and plasmids. *Escherichia coli* DH5α cells and *E. coli* XL1-Blue MR cells were purchased from BRL/Life Technologies, Gaithersburg, Md., and Stratagene Cloning Systems, La Jolla, Calif., respectively. *E. coli* S17-1 was obtained from Philippe Mazodier, Institut Pasteur, Paris, France. Bacteriophage T7 was obtained from Phil Youngman, University of Georgia, Athens. Plasmid pGEM-3Zf was purchased from Promega Corporation, Madison, Wis. Plasmids pUC19 (35), pKC1139 (6), and pNJ1 (31) have been described previously.

Nucleic acid isolation and manipulation. Plasmid isolation from *E. coli* was performed by using a Qiagen Midi kit and Qiaprep Spin Plasmid kit (Qiagen GmbH, Hilden, Germany). DNA cloning and manipulations were performed by using standard procedures (27). Restriction enzymes, T4 DNA ligase, calf intestinal alkaline phosphatase (CIAP), and the Klenow fragment of DNA polymerase I were purchased from BRL/Life Technologies. Hybond-N membranes were obtained from Amersham Corporation, Arlington Heights, Ill.

* Corresponding author. Mailing address: Abbott Laboratories, D-47P AP-9A, 100 Abbott Park Road, Abbott Park, IL 60064. Phone: (847) 938-7389. Fax: (847) 938-3403. E-mail: stassid@ugene1.pprd .abbott.com.
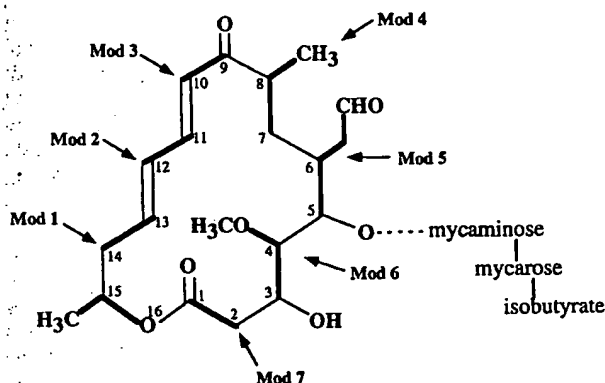
FIG. 1. Structure of niddamycin. The atoms of the macrolide ring are numbered. The thick lines indicate carbon atoms contributed by the indicated modules (Mod).

Construction of a *Streptomyces caelestis* genomic library. *S. caelestis* DNA was prepared by partially digesting 5 μg of DNA with *Sau*IIIA and size selecting fragments of around 40 kb by electrophoresis through a 0.4% agarose gel. Cosmid pNJ1 arms were prepared by digestion of the vector with *Eco*RI, dephosphorylation with CIAP, and then digestion with *Bgl*II to generate one arm and digestion of the cosmid with *Hind*III, dephosphorylation with CIAP, and then digestion with *Bgl*II to generate the other. Ligation of 1 μg of cosmid arms to 1 μg of the size-selected *S. caelestis* DNA was in a 20-μl volume. Two microliters of this ligation mix was packaged by using GigapackII XL (Stratagene) as instructed by the manufacturer, and *E. coli* XL1-Blue MR cells were hosts for infection. Individual colonies were transferred to 30 96-well plates (Costar, Cambridge, Mass.).

S. caelestis PKS probe. A PKS-specific probe was generated by PCR amplification of *S. caelestis* genomic DNA, with degenerate primers designed from KS and AT sequences in the GenBank/EMBL database. The KS-specific oligonucleotide (5'-CGGTSAAGTCSAACATCGG-3') and the AT-specific oligonucleotide (5'-GCRATCTCRCCCTGCGARTG-3') were used in a PCR mixture containing ThermoPol reaction buffer (New England Biolabs, Beverly, Mass.), 0.2 mM deoxynucleoside triphosphate mixture, 0.5 μg of genomic DNA, and 100 pmol of the primers in a volume of 98 μl. After incubation at 96°C for 2 min, 2 U of Vent DNA polymerase (New England Biolabs) was added, and the mixture was cycled 25 times in a Perkin-Elmer Cetus 9600 thermocycler at 96°C for 30 s, 50°C for 2 min, and 72°C for 4 min, followed by a final incubation at 72°C for 15 min. The reaction generated a 900-bp DNA fragment which was gel purified from a 1% agarose gel by using a Prep-A-Gene kit (Bio-Rad Laboratories, Hercules, Calif.).

Hybridization of the PKS probe. The *S. caelestis* library was transferred from 96-well plates to Hybond-N membranes for hybridization. The membranes were soaked in prehybridization solution (27) for at least 2 h at 65°C. About 50 ng of the KS/AT probe was labeled with [α-³³P]dCTP (DuPont, NEN Research Products), using the Megaprime DNA labeling system (Amersham) as instructed by the manufacturer. The radiolabeled probe was added to the prehybridization solution and incubated at 65°C for 16 to 20 h. The membranes were then washed twice in 1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate at room temperature and once in 0.1× SSC–0.1% sodium dodecyl sulfate at 65°C for 20 min. The membranes were dried and placed on 35-by 43-cm phosphor screens (Molecular Dynamics, Sunnyvale, Calif.) for 2 to 4 h. Imaging of the screens was on a Molecular Dynamics PhosphorImager 425.

DNA sequencing. Cosmids that hybridized to the KS/AT probe were digested with *Sst*I, and the resulting fragments were subcloned into vector pGEM-3Zf or pUC19. *Sst*I fragments that were larger than 8 kb were further digested with *Sma*I and subcloned to generate plasmids that would be more suitable for sequencing. Two methods were used for DNA sequencing. One method used [α-³³P]dCTP (DuPont, NEN Research Products) and the fmole DNA cycle sequencing system (Promega). Samples were run on 6% polyacrylamide–8 M urea gels, using a Sequi-Gen II sequencing apparatus (Bio-Rad). The second method used an ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer Corporation), and the samples were run on 4.75% polyacrylamide–8.3 M urea gels on an Applied Biosystems 373 sequencer. All subclones were sequenced on both strands by using multicloning site primers to initiate the sequences and internal primers to extend the sequences.

DNA sequence data was analyzed by using the Wisconsin sequence analysis package programs (Genetics Computer Group, Madison, Wis.) (8). Database searches to identify homologs to the deduced amino acid sequences were performed with BLAST (1).

Conjugation. *S. caelestis* recipient cells were prepared by inoculating 30 ml of SGGP with 25 μl of spores and incubating the culture overnight at 30°C on a rotary shaker. The culture was then centrifuged at a relative centrifugal force of 2,190 for 15 min. The pelleted cells were washed once in SGGP, recentrifuged, and resuspended in 2 ml of SGGP. *E. coli* donor cells, S17-1 and S17-1(pSK7595), were prepared by growing overnight cultures in 2 ml of LB (with apramycin [50 μg/ml] where appropriate) at 37°C on a rotary shaker. Two hundred microliters of the *E. coli* cultures was inoculated into 30 ml of fresh medium and grown for about 3 h (to mid-log phase) at 37°C. The cells were pelleted by centrifugation for 20 min and washed once in the same medium. After a final centrifugation, the pellets were resuspended in 300 μl of LB. Nucleopore membranes (25 mm, 0.2 μm; VWR) were placed onto an AS-1 agar plate and spotted with 50 μl each of donor and recipient cells. The plates were incubated at 30°C for 3 h, and then each membrane was placed into a 50-ml conical tube containing 5 ml of SGGP and vortexed for 30 s to dislodge the cells. The cells were centrifuged for 20 min, and the pellets were resuspended in 100 μl of SGGP. Cells were then plated onto AS-1 plates and incubated overnight at 30°C. The next day, the plates were overlaid with 2 ml of distilled H₂O containing 30 μl of a 50-mg/ml solution of apramycin and 150 μl of T7 bacteriophage (~10¹⁰ PFU) and returned to 30°C until transconjugants appeared (about 7 days).

Metabolite analysis. Culture supernatants of cells grown for 3 days at 30°C in SCM medium were adjusted to pH 9.0 with NH₄OH and extracted two times with equal volumes of ethyl acetate. The organic phases were pooled and concentrated. Extracts were evaluated on silica gel thin-layer chromatography (TLC) plates (60F-254; Merck), which were developed with isopropyl ether-methanol-NH₄OH (75:35:2). Compounds were visualized by spraying the plates with anisaldehyde-sulfuric acid-ethanol (1:1:9) and heating the plates with a hot air gun until color developed. Biological activity was assessed by placing unstained TLC plates onto AM11 agar plates seeded with *Staphylococcus aureus* as the indicator strain and incubating the plates overnight at 37°C to develop zones of inhibition.

Nucleotide sequence accession number. The GenBank accession number for the niddamycin PKS and portions of two flanking genes is AF016585.

RESULTS

Cloning and sequencing of the niddamycin PKS cluster in *S. caelestis*. A strategy was devised to isolate the PKS genes of the niddamycin biosynthetic pathway by using sequence conservation in the KS and AT regions encoded by previously sequenced PKS genes (5, 10). A pair of degenerate primers spanning conserved regions was designed from sequences deposited in GenBank (see Materials and Methods) with the expectation of amplifying, from *S. caelestis* chromosomal DNA, most or all of the genes for the seven KS/AT regions predicted to be found in the niddamycin PKS. The product of the PCR ran as a 900-bp fragment on an agarose gel (data not shown) and was subsequently cloned into M13 for sequencing. Seven clones were analyzed, and all contained the same sequence, which was later found to correspond to the KS/AT region in module 5 of the niddamycin cluster (see below). This PCR product was used to probe chromosomal DNA to determine whether it had sufficient homology to hybridize to the remaining KS/AT domains of the niddamycin biosynthetic cluster and therefore could be used to identify cosmids in a library of *S. caelestis* DNA which encoded the remainder of the pathway. When DNA from *S. caelestis* was digested with *Sst*I and probed with the PCR fragment, at least seven hybridizing bands were visible (data not shown). A second Southern hybridization, but with a 1.45-kb *Sst*I-*Msc*I fragment (coordinates 10793 to 12244; GenBank accession number M63677) encoding part of the KS domain from module 5 of the erythromycin PKS, gave the same banding pattern, suggesting that these bands may represent all of the type 1 PKS genes present in *S. caelestis*.

To construct the *S. caelestis* library, a partial *Sau*IIIA digest of genomic DNA which had been size selected for 30-kb fragments was ligated to cosmid pNJ1 arms and propagated in *E. coli* XL-1 Blue MR cells. The library of 2,880 clones was screened by colony blotting with the *S. caelestis* KS/AT probe described above. Nineteen clones gave a strong hybridization signal, and restriction digests of the positive clones were ana-
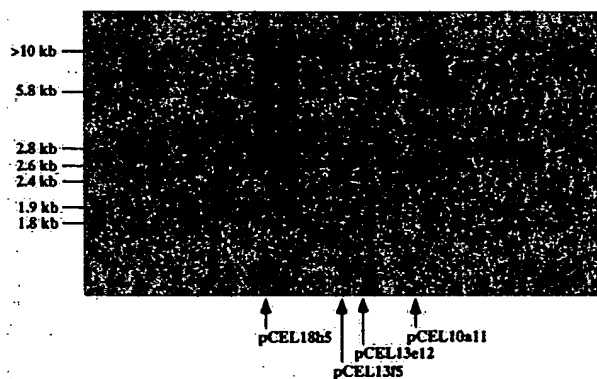
FIG. 2. Southern analysis of positively hybridizing cosmids. Nineteen cosmids that hybridized with the KS/AT probe were digested with *Sst*I and analyzed by Southern hybridization using a KS/AT probe generated from *S. caelestis* DNA. Three of the cosmids did not cross-react (i.e., were false positives). Lanes containing cosmid DNA specifically referred to in the text are labeled.

lyzed by Southern hybridization. Banding patterns of *Sst*I-digested cosmid DNAs probed with the same DNA indicated that 16 of the 19 clones which were positive in colony blots were also positive by Southern analysis and that 15 of these clones contained overlapping inserts (Fig. 2). Three of the clones contained six hybridizing bands, while the other 12 overlapping clones contained from two to five of the same hybridizing bands. Clone pCEL18h5, with six hybridizing bands, was chosen for sequence analysis. Two other clones, pCEL13f5 and pCEL13e12, not only shared a large hybridizing fragment (>10 kb) with pCEL18h5 but also contained a unique 2.4-kb fragment, indicating that these cosmids could contain PKS sequences not present in pCEL18h5. Another related clone, pCEL10a11, contained the 2.4-kb fragment but not the larger fragment, and so it was not analyzed further. The

large *Sst*I fragment was isolated from clones pCEL13f5 and pCEL18h5 and further digested with *Sma*I for Southern analysis. Probing indicated that both clones contained hybridizing bands of 2.1, 2.2, and 3.2 kb, but clone pCEL13f5 also contained a 1.9-kb hybridizing band not found in clone pCEL18h5. This was further evidence that clone pCEL13f5 shared DNA regions with clone pCEL18h5 but also contained neighboring PKS sequences. For this reason, clone pCEL13f5 was also analyzed along with clone pCEL18h5.

Restriction fragments from cosmids pCEL18h5 and pCEL13f5 (Fig. 3) were subcloned into pUC19 or pGEM-3Zf and sequenced as described in Materials and Methods. Orientation of the subcloned fragments was determined by using the parent cosmid as the template for primers that annealed to the 5' and 3' ends of each insert, generating upstream and downstream sequences. These could then be aligned with sequences of the individual subclones to order them within the cosmid. Each cosmid was found to contain approximately 30 kb of insert DNA, with around 10 kb of overlapping sequence. In all, 41.1 kb of DNA was sequenced.

Organization of the niddamycin gene cluster. DNA sequence data obtained from cosmids pCEL18h5 and pCEL13f5 were analyzed for open reading frames (ORFs) by using a *Streptomyces* codon usage table (33). Five large ORFs (ORFs 1 to 5) spanning 40,012 bp of DNA were revealed (see below). These large ORFs were bounded by smaller, partially sequenced ORFs (ORFs 6 and 7) located at the 5' and 3' ends of ORFs 1 to 5, respectively (Fig. 3). The deduced N-terminal 132-amino-acid sequence encoded by ORF 6 was found, by BLAST search, to have 61% identity to the N-terminal region of the product of the *tylI* gene, which encodes a cytochrome P450 hydroxylase in the biosynthetic pathway of the 16-membered macrolide tylosin (19). Like the gene containing ORF 6, *tylI* in *Streptomyces fradiae* is also located immediately 5' to, and reads in the opposite direction from, the genes encoding the PKS for tylosin production. The deduced N-terminal sequence encoded by ORF 7 (157 amino acids) was found to

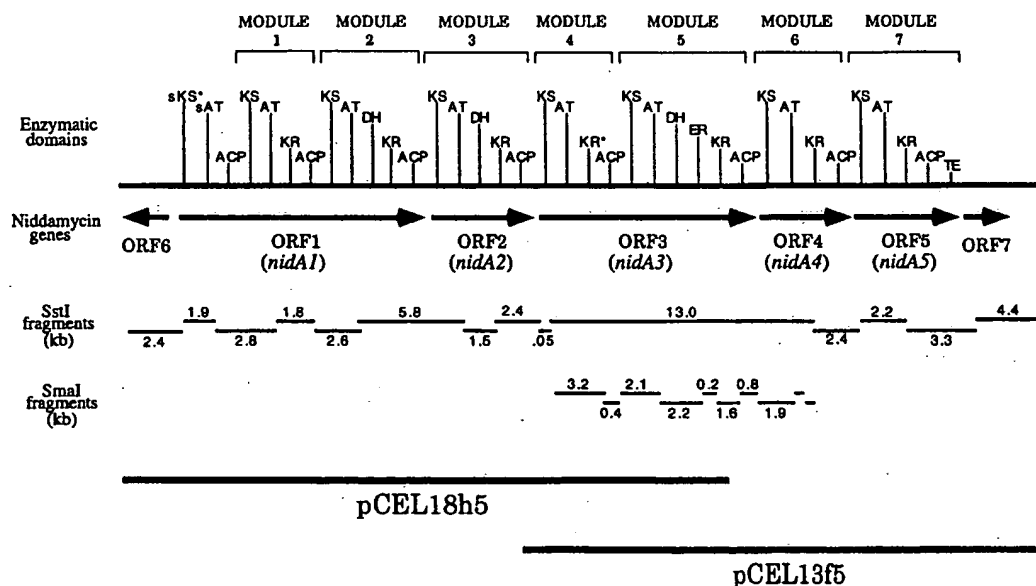FIG. 3. Organization of the niddamycin PKS genes. Cosmids pCEL18h5 and pCEL13f5 were isolated from a genomic library of *S. caelestis* DNA. DNA sequence was obtained from subcloned restriction fragments indicated by narrow lines and from cosmid DNA to link the restriction fragments. ORFs, enzymatic domains, and modules were identified from DNA sequence analysis. Domains labeled with asterisks are probably inactive.

TABLE 1. Niddamycin PKS sequence coordinates

ORF no.	Nucleotide coordinates	No. of amino acids	Homology
6	1–396	132	P450 (N-terminal portion)
1	627–13646	4340	PKS (loading domain, modules 1 and 2)
	690–1904	405	KS* (inactive)
	2235–3242	336	ATs
	3312–3572	87	ACPs
	3645–4916	424	KS
	5214–6215	334	AT
	6900–7442	181	KR
	7695–7955	87	ACP
	8025–9296	424	KS
	9594–10595	334	AT
	10650–11225	192	DH
	12219–12770	184	KR
	13068–13325	86	ACP
2	13693–19209	1839	PKS (module 3)
	13795–15066	424	KS
	15379–16380	334	AT
	16423–16992	190	DH
	17839–18393	185	KR
	18676–18933	86	ACP
3	19269–30230	3654	PKS (modules 4 and 5)
	19374–20645	424	KS
	20973–22007	345	AT
	22530–23057	176	KR* (inactive)
	23322–23579	86	ACP
	23646–24926	427	KS
	25293–26315	341	AT
	26355–26909	185	DH
	27789–28835	349	ER
	28842–29393	184	KR
	29679–29936	86	ACP
4	30294–35000	1569	PKS (module 6)
	30423–31691	423	KS
	32013–33038	342	AT
	33786–34337	184	KR
	34593–34853	87	ACP
5	35000–40636	1879	PKS (module 7, TE)
	35120–36388	423	KS
	36692–37723	344	AT
	38633–39184	184	KR
	39455–39712	86	ACP
	39824–40636	271	TE
7	40626–41097	157	N-Methyltransferase (N-terminal portion)

have 68% identity to the product of ORF 3*, an N-methyltransferase of the tylM1-encoded region of the tylosin biosynthetic cluster (12). Like ORF 7, ORF 3* is downstream of and reads in the same direction as the genes encoding PKS; however, it is not immediately adjacent to the genes encoding PKS. Based on sequence alignments and distance from potential ribosome binding sites (RBSs) (30), ORFs 6 and 7 appear to use ATG as the start codon (Table 1), with the putative RBSs being GGAGA and GGAGG, respectively. In addition, ORF 7 may be translationally coupled to ORF 5, since the putative ATG start codon of ORF 7 begins 11 bases upstream of the TGA stop codon of ORF 5.

The larger ORFs (ORFs 1 to 5) were found to encode a PKS comprising proteins with molecular masses of 458, 194, 379,

167, and 198 kDa, respectively. The genes corresponding to these ORFs were designated nidA1 through nidA5, respectively. nidA1, nidA2, and nidA5 appear to begin with ATG codons (Table 1) and have the putative RBSs GGAGG, GGGGG, and GGAG, respectively. nidA3 and nidA4 appear to begin with GTG codons, with the RBSs being GGAGA and GGAAG, respectively. The enzymatic motifs encoded by these genes (i.e., KS, AT, DH, ER, KR, and ACP) were delineated by BLAST searches which identified homology to other PKS genes deposited in the EMBL, GenBank, and Swissprot databases. As is characteristic of type I PKSs, the motifs within each ORF product are organized into modules. Seven modules, corresponding to the seven condensation reactions predicted to be required for niddamycin biosynthesis, were found in the PKS; NidA1 and NidA3 contain two modules each, whereas NidA2, NidA4, and NidA5 each contain one module (Fig. 3). In addition, the sets of motifs contained within each module were also consistent with the extent of β-carbon processing at the corresponding positions of the niddamycin ring (Fig. 1 and 3). Besides the KS, AT, and ACP domains found in all modules, modules 2 and 3 have the KR and DH domains required for the formation of the double bonds found between carbons 12 and 13 and carbons 10 and 11, respectively. Module 5 has a full complement of domains (KS, AT, DH, ER, KR, and ACP) to produce the methylene group at C-7. Modules 4, 6, and 7 each have KR domains which would result in hydroxyl groups at C-9, C-5, and C-3, respectively. However, sequence analysis of the KR in module 4 indicates that it is inactive (see below), resulting in the keto group at C-9. A TE domain with the conserved GxSxG and GdH motifs (10) was found at the carboxy-terminal end of NidG5 (module 7), identifying the end of the PKS cluster.

In contrast to the erythromycin PKS, where the N-terminal enzymatic motifs are a loading AT and ACP, the loading AT and ACP of the niddamycin PKS follow a KS domain which is not predicted to be necessary for the initiation of polyketide synthesis. Homology of this motif with the other KS domains in the cluster ranges from only 45 to 47% identity, whereas homologies among the KS domains of modules 1 to 7 range from 70 to 92% identity. More importantly, sequence analysis indicates that this KS is enzymatically nonfunctional, as a critical cysteine residue in the motif TVDTGCSSSLV, which is highly conserved among KS domains (3, 10), is replaced by a glutamine residue at position 173, forming a KS^Q domain (Fig. 4). Interestingly, it has recently been reported that KS^Q domains are also found at the N termini of the PKSs which synthesize the other 16-membered macrolides, carbomycin, spiramycin, and tylosin (17). The function of this motif, if any, remains to be determined.

All of the remaining seven KS domains of this cluster retain the conserved active-site cysteine residue. In addition, two highly conserved $His^{308/347}$ residues (3, 11) are present in all of the niddamycin KS domains, including the unusual starter KS (Fig. 4).

The niddamycin AT domains show more sequence variability than the KS domains. For example, AT2 and AT6 only have 30% amino acid identity, whereas AT2 and AT3 have 95% amino acid identity. It has been demonstrated that mATs and mmATs fall into distinct families based on amino acid sequence and can be distinguished by conserved motifs (28). As a result, it may be possible to predict AT substrate specificity on the basis of primary sequence. A PILEUP/DENDRO GRAM analysis comparing the niddamycin ATs with those from the rapamycin and erythromycin PKS clusters is shown in Fig. 5. This analysis demonstrates that the loading AT, as well as those of modules 1, 2, 3, and 7, fall into the group of ATs

```
KS            171              305           344
  NidKS1    TACSSSLVAL-----VEAHGTGTAL-----NIGHTQAAAG
  NidKS2    TACSSSLVAL-----VEAHGTGTPL-----NIGHSQAAAG
  NidKS3    TACSSSLVAL-----VEAHGTGTPL-----NIGHTQAAAG
  NidKS4    TACSSSLVAL-----VEAHGTGTPL-----NIGHTQAAAG
  NidKS7    TACSSSLVAL-----VEAHGTGTPL-----NIGHTQAAAG
  NidKS6    TACSSSLVAL-----VEAHGTGTPL-----NIGHTQAAAG
  NidKS5    TACSASLTAL-----VEAHGTGTHL-----NIGHTQAAAG
  NidKSs*   TGQSSSLVAV-----VELHGTGTPA-----NIGHLEGAAG
                    •

DH             34
  NidDH2    VSLAAHPWLA DHTVAGAVLP PGTAFADLLL
  NidDH5    IGVDEFPWMA DHTLLGAVLL PGAAFADLAL
  NidDH3    LSARALPWLA DHLVWDRGVA PGTAVLEIVL
                    •

KR             1
  NidKR4*   GTVLVVGDST TVPARLLRSL LD.DGAERVV
  NidKR6    GTILITGGTG ALGTHIATWL AHK.GAKHLI
  NidKR7    GTILITGGTG ALGTHIATWL AHK.GAKHLI
  NidKR3    GTVLITGGTG ALAAVTARHL VARHGARHLL
  NidKR5    GTVLITGGTG TIGSRIARHL VTRHGVRHLL
  NidKR2    GTVLLTGCTG ALGRRVAVHL ARRHGVRRML
  NidKR1    GTVLITGGTG ALQSQVARRL AL.AGAPHLL

ER            154
  NidER5    GLVALAGLRR GETVLVHAAA GGVGMAAVQI ARHLGAQVLA
  EryER4    ALHDLAGLRA GQSVLIHAAA GGVGMAAVAL ARRAGAEVLA

ACP           35
  NidACP4   STRAFRELGF ESLTAVELRG RLAEATGLTL
  NidACP7   ARASFRDLGF DSLAAVSLRD GLAEATGLEL
  NidACP6   HDVAFKDLGF DSLAAVKMRT RLRETTGLDL
  NidACP3   AERSFKDAGF DSLTAVDLRN RLNARTGLRL
  NidACP5   SDRAFRBAGF DSLTAVELRN RLAAATGLRL
  NidACP2   PERPFHEIGF DSLATLELRN RLGRLVGLRL
  NidACP1   AETAFRAGGF DSLTLVELRN RLASATGLRL
  NidACPs   LDRPFTSQGL DSMTAVELAG LLGTAAGVAL
                    •

TE            101
  NidTE     PFALAGHSAG ANVAYALAAY--------DVPGNHPTML
  EryTE     PFVVAGHSAG ALMAYALATE--------AVPGDHPTMV
                    •
```

FIG. 4. Conserved active sites of the niddamycin PKS domains. Deduced protein sequences were aligned by using the PILEUP/PRETTY program. Only the regions containing the proposed active sites are shown. The numbers correspond to the amino acid position from the first residue in the first motif listed. The boldfaced amino acids are conserved residues or motifs important for the function of each domain. Active-site residues are marked with asterisks. Motifs important for functioning of the KR and ER domains are underlined. For comparative purposes, the erythromycin ER from module 4 and the erythromycin TE are included.

that use malonyl-CoA as the substrate. Based on the structure of niddamycin (Fig. 1), the acetate-derived methyl group at C-15 and the lack of methyl groups at C-14, C-12, C-10, and C-2 are consistent with those ATs being malonyl-CoA specific. On the other hand, the ATs of modules 4, 5, and 6 fall into the methylmalonyl-specific group. The methyl group at C-8 is consistent with AT4 being an mmAT. AT5, however, presumably uses ethylmalonyl-CoA as a substrate, resulting in the ethyl side chain at C-6. The fact that it does not segregate into a unique group in the dendrogram indicates that methylmalonyl and ethylmalonyl ATs may be relatively similar. The AT domain in module 6 also falls into the methylmalonyl family, but its activity results in a hydroxyl or perhaps a methoxy group at C-4. The biosynthesis of leucomycin, a 16-membered macrolide which contains a methoxy group at C-4, was examined by feeding the producing organism, *Streptoverticillium kitasatoensis*, 2-^{13}C-labeled precursors (23, 24). Labeled malonate was not incorporated at carbons 3 and 4, indicating that malonyl-CoA was not a substrate for the corresponding AT. Labeled glycerol, however, was incorporated at C-4, suggesting that glycolate may be the substrate for the AT of the module corresponding to the C-4 position of leucomycin. It is conceivable
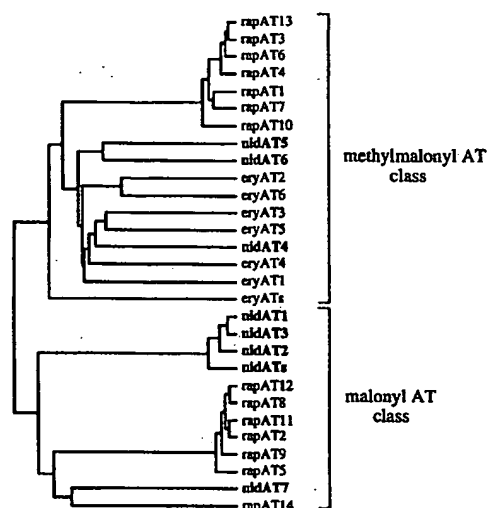

FIG. 5. Dendrogram analysis of AT domains. The dendrogram shows the sequence similarity of AT domains from the niddamycin, erythromycin, and rapamycin PKS clusters. The niddamycin AT domains are boldfaced. The ATs are grouped according to the substrate specificities of the rapamycin and erythromycin ATs (malonyl-CoA or methylmalonyl-CoA).

that the AT in module 6 of the niddamycin pathway uses the same or a similar substrate.

A closer examination of the amino acid sequence (Fig. 6) of the niddamycin ATs shows that the active-site residues Gln[6], Ser[91] (of the GHSXG motif where Ser is involved in the formation of the acyl-enzyme intermediate), Arg[116], and His[192]

```
MALONYL AT CLASS

           1               58                      89
  NidAT1  LLPTGQGAQH---RTHYT.......QTALYRTL..........GHSVGEI---
  NidAT3  LLPTGQGAQH---RTHYT.......QTALYRTL..........GHSVGEI---
  NidAT2  LLPTGQGAQH---RTHYT.......QTALYRTL..........GHSVGEI---
  NidATs  LLPTGQGAQH---RTHYT.......QTALYRTL..........GHSVGEI---
  NidAT7  FLPSGQGAQR---RTHYT.......•TALfR1f..........GHSIGGL---
                    •
  con. acet.        FTGYA.......Q.A.FGLL..........GHSXG

           113             187
  NidAT1  ITARAHLMGQ---VSHAFHSALM
  NidAT3  ITARAHLMGQ---VSHAFHSALM
  NidAT2  ITARAHLMGQ---VSHAFHSALM
  NidATs  ITARAHVMGQ---VSHAFHSALM
  NidAT7  VAARGRLMQR---VSHAFHSPHM
                    •       •

METHYLMALONYL AT CLASS

           1               58                      89
  NidAT5  FVFPGQGSQW---RVDVV.......MVSLAAVW..........GHSQGEI---
  NidAT6  FVFPGQGAQW---RaDVV.......MVSLAAlW..........GHSQGEV---
  NidAT4  FVFPGQGAQW---RVDVV.......VVaiAAnW..........GHSQGEI---
                    •                              •
  con. prop.        KVDVV.......M.S.AAhW..........GHSQG

           113             190
  NidAT5  VALRSRAWLG---VDtAGHSPQV
  NidAT6  IAGRSRLWGR---VDfAGHSPQV
  NidAT4  VALRSQLIAR---VDyAHSRHV
                    •       •
```

FIG. 6. Alignment of the sequences of conserved motifs found in AT domains. Sequences were aligned by using the PILEUP/PRETTY program. Boldfaced amino acids are conserved residues or motifs important for the function of AT domains. Active-site residues are marked with asterisks. Proposed consensus sequences for motifs of ATs specific for acetate and propionate (13) are indicated.

are present in all mATs and mmATs (3, 10). In addition, motifs which have been proposed to distinguish mATs from mmATs (13) seem to be more divergent in the niddamycin cluster (Fig. 6, residues 58 to 77); however, these same stretches of amino acids, though they have different sequences, still seem to be specific for mATs and mmATs within the niddamycin sequence.

KR domains were found in all seven modules. Homologies between the KR domains ranged from 31% amino acid identity between KR4 and KR6 to 99% amino acid identity between KR6 and KR7. It has been observed that modules corresponding to unreduced β-carbonyls in the polyketide chains of erythromycin (9), rapamycin (3), and spiramycin (17) still contain KR motifs with, in most cases, obvious inactivating mutations. An inactive KR in module 4 would be consistent with the presence of the keto group at C-9 of niddamycin and would be another example of this type of organization. The consensus NADP(H) binding site of a KR is GxGxxGxxxA (29). The corresponding region of the KR4 domain of the niddamycin cluster is DxTxxPxxxL (Fig. 4), indicating that the KR is nonfunctional. All of the other KR domains in the cluster contain the expected NADP(H) binding motif with the exception of KR2. This KR is predicted to be active; however, a Cys residue is present at the first position of the NADP(H) binding site rather than a Gly. How this amino acid change would affect KR activity is uncertain.

DH domains were identified in modules 2, 3, and 5. Hydroxyl groups formed by KRs at C-13 and C-11 are reduced by DHs to form the double bonds between carbons 12 and 13 and carbons 10 and 11, corresponding to modules 2 and 3, respectively. Formation of the methylene group at C-7 (corresponding to module 5) also requires DH activity. DH2 and DH3 have 39% amino acid identity, DH2 and DH5 have 45% identity, and DH3 and DH5 have 38% identity. The highly conserved residues His[45], Gly[49], and Pro[54] (5, 10) were present in DH2 and DH5, but, surprisingly, in DH3 there is an Asp in place of the Gly residue (Fig. 4). Based on the niddamycin structure, this DH domain should be active, and so this divergence from consensus may not adversely affect the activity of the enzyme.

The only ER domain present in the PKS cluster is in module 5. An amino acid sequence comparison of ER5 with ER domains from the products of the erythromycin and rapamycin PKS genes (Fig. 4) reveals the presence of the putative NADP(H) binding motif LxHxg(a)xGGVG (2, 29, 32).

ACP domains were identified in each of the modules and after the loading AT, as expected. Homologies between the ACP domains range from 32 to 67% amino acid identity. The pantotheine-binding Ser[46] residue (10) in the GFDSL motif was present in all of the ACP domains (Fig. 4).

Gene disruption of the niddamycin PKS cluster. A gene disruption experiment was performed with *S. caelestis* to demonstrate that the PKS genes isolated from the genomic library were involved in niddamycin biosynthesis. A 1.5-kb SstI fragment containing AT- and DH-encoding sequences from the coding regions for module 3 (Fig. 2) was isolated from cosmid pCEL18h5, treated with Klenow fragment, and cloned into the EcoRV site of pKC1139 to generate pSK7595 (Fig. 7). This plasmid contains an apramycin resistance gene for selection, oriT for conjugal transfer, and both *E. coli* and *Streptomyces* origins of replication. The *Streptomyces* origin of replication from *Streptomyces ghanaensis* (21) functions only at temperatures below 34°C, and so under selective pressure at a nonpermissive temperature, the plasmid must integrate into the chromosome through homologous recombination to be maintained. Plasmid pSK7595 was first transformed into *E. coli* S17-1, the donor strain carrying the conjugation locus RP4. To transfer the
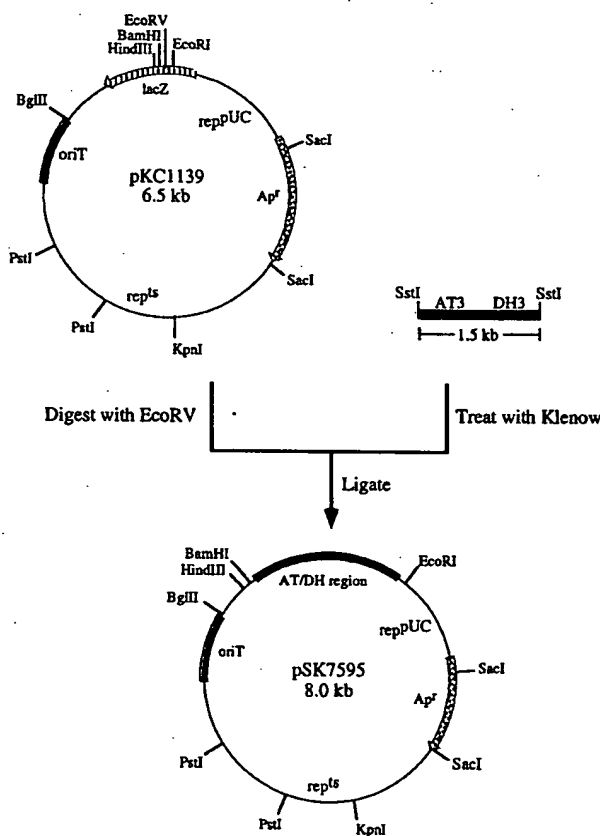

FIG. 7. Construction of pSK7595. The vector pKC1139 was used to construct the conjugation vector pSK7595. The oriT gene allows for conjugal transfer from *E. coli* strains containing the RP4 locus. The apramycin resistance gene (Ap^r) allows for selection of the plasmid. The temperature-sensitive *Streptomyces* origin of replication (rep^{ts}) and the *E. coli* origin of replication (rep^{pUC}) are shown. The 1.5-kb SstI fragment containing AT- and DH-encoding sequences was cloned into the EcoRV site of pKC1139 (black bar).

plasmid to the niddamycin-producing strain, S17-1 transformants were incubated with *S. caelestis* cells as described in Materials and Methods. As a negative control, S17-1 cells without plasmid were also incubated with *S. caelestis* cells. The cultures were challenged with apramycin to eliminate nonplasmid-containing *S. caelestis* cells. Nalidixic acid (50 μg/ml) is routinely used in *Streptomyces-E. coli* conjugation experiments to eliminate the *E. coli* donor cells; however, *S. caelestis* was found to be sensitive to this antibiotic. Therefore, the conjugation cultures were challenged with bacteriophage T7 in order to remove the *E. coli* cells. Cultures were incubated for 7 days at 30°C, and in each of two separate experiments, only one *S. caelestis* colony was recovered.

One putative transconjugant, *S. caelestis*(pSK7595), was grown at 37°C in apramycin-containing medium to force integration of the temperature-sensitive plasmid. Southern analysis of genomic DNA indicated that the plasmid was present and had integrated into the appropriate location in the chromosome (Fig. 8). This strain was designated *S. caelestis* nid:: SK7595. Surprisingly, the original transconjugant isolated at 30°C and never grown at a nonpermissive temperature also showed the pattern expected for integrated plasmid, indicating that this plasmid was not replicating at the permissive temperature in *S. caelestis*. This may also explain the extremely low
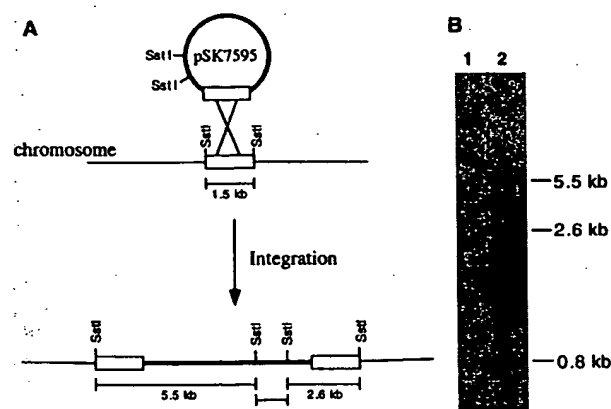
A

B

FIG. 8. Disruption of the niddamycin PKS locus. (A) Diagrammatic representation of the integration of pSK7595 into the *S. caelestis* chromosome. The region of homology between the plasmid and chromosome is represented by an open box. pKC1139 DNA is represented by a heavy line. *Sst*I sites are indicated. (B) Southern analysis of *Sst*I-digested genomic DNA probed with vector pKC1139. Lane 1, *S. caelestis* NRRL-2821 (wild type); lane 2, *S. caelestis*(pSK7595). DNA fragment sizes based on migration of a 1-kb ladder (BRL) on the same gel are indicated at the right.

frequency of conjugation since transconjugants could be selected only if a rare integration event took place before the plasmid was lost.

Spores of *S. caelestis* nid::SK7595 were plated onto R3M medium containing apramycin to generate single colonies. Three of the isolates were examined for the production of niddamycin. Cells were grown in apramycin-containing medium at 30°C (data not shown) and 37°C for 4 days, and supernatants were examined by TLC and TLC-bioautography (Fig. 9). None of the *S. caelestis* nid::SK7595 isolates, grown at either temperature, produced a spot on the TLC plate at the R_f of niddamycin. In addition, no bioactivity was observed in the

A   **B**

FIG. 9. Niddamycin production in *S. caelestis*(pSK7595). (A) TLC analysis of ethyl acetate extractions of culture supernatants. Lanes: wt (wild type), *S. caelestis* NRRL-2821; 1 to 3, isolates of *S. caelestis*(pSK7595). nid std (niddamycin standard), 10 μg of niddamycin. The R_f of niddamycin (nid) is indicated by the arrow. (B) TLC-bioautography analysis of ethyl acetate extractions of culture supernatants. The lane assignments and niddamycin location are identified as in panel A. Dark spots are zones of inhibition of *S. aureus* due to the presence of niddamycin.

TLC-bioautography assay, indicating that the niddamycin PKS cluster had been disrupted and that the cloned genes were from the niddamycin pathway.

DISCUSSION

We have reported the isolation and characterization of the PKS genes involved in the biosynthesis of the macrolide antibiotic niddamycin. The initial approach to the isolation of this cluster was to PCR amplify AT regions based on amino acid sequences which are found to be highly conserved among PKSs. We generated primers corresponding to regions in the KS and AT domains with the expectation of amplifying DNA corresponding to portions of all seven domains hypothesized to be present in the cluster. (The subsequent discovery that a KS^Q domain preceding the loading AT brings the number of potentially amplifiable regions to 8.) The PCR resulted in the coding region for only one AT, that from module 5, being amplified. On reexamination of the homology of the niddamycin nucleotide sequences to the primers, it was found that there was only one mismatch near the 5′ end of the KS primer among the eight niddamycin KS nucleotide sequences but a much higher frequency of mismatches at the 3′ end of the AT primer. For example, there were six mismatches each for five of the niddamycin AT nucleotide sequences, three mismatches for one and one for another. The coding region for module 5, which was successfully amplified, had no mismatches. Therefore, this PCR approach proved fruitful for generating a niddamycin-specific probe to isolate the remaining coding regions of the cluster but not for simultaneously amplifying the coding regions for multiple domains.

Probing the *S. caelestis* genome with either the PCR-generated probe or a probe from a conserved region of the erythromycin PKS generated the same hybridization patterns, indicating that these were the only type 1 PKS sequences present in the strain. It has been observed that some strains of *Streptomyces*, such as *Streptomyces hygroscopicus* 29253, contain multiple PKS clusters (18, 26), which can complicate isolation of a particular cluster by hybridization. This does not appear to be the case in *S. caelestis*. When the PCR-generated probe was used for Southern analysis of *Sst*I-digested *S. caelestis* DNA, we identified seven cross-hybridizing bands which corresponded to seven of the *Sst*I fragments encoding the niddamycin PKS.

Sequence analysis of the isolated DNA fragments revealed a PKS cluster comprising seven modules as expected for synthesis of niddamycin. The modules are encoded by five ORFs, with two modules each encoded by ORFs 1 and 3 and one each encoded by ORFs 2, 4, and 5. The organization of the enzymatic domains contained within each module is also consistent with what would be expected for synthesis of the macrolide ring of niddamycin. An unexpected KS domain (KS^Q) with a glutamine replacing a conserved cysteine residue was found at the amino terminus of the PKS. The function of this domain, if any, remains unclear. Platenolide A is the polyketide backbone of the macrolide antibiotic spiramycin and is identical in structure to the polyketide backbone of niddamycin. A genetic map of the spiramycin PKS was recently published (17), and the organization and domain content of the modules, including the unusual KS^Q, are the same as those found in this study for the niddamycin cluster. The degree of genetic relatedness of these clusters awaits the release of the spiramycin PKS nucleotide sequence.

Previous work has suggested that AT sequences cluster into families based on substrate specificity (13, 28). As predicted from the structure of niddamycin, the AT domains of modules
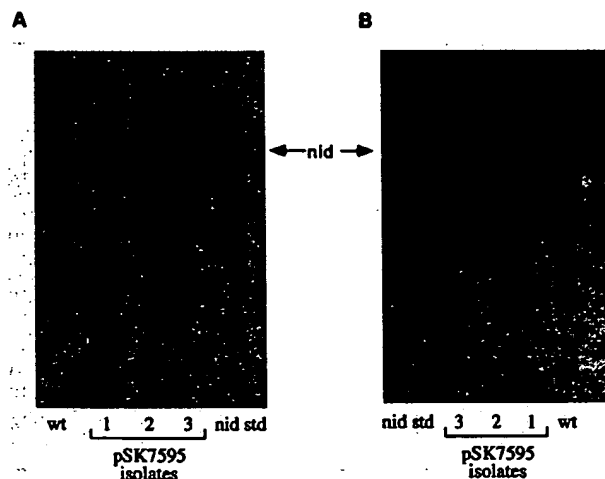
1, 2, 3, and 7 and the loading AT fall into the class of ATs which utilize malonate. The remaining ATs fall into the methylmalonate class. The methyl group at C-8 is consistent with AT4 being an mmAT. AT5 and AT6, however, are predicted to utilize, respectively, ethylmalonyl-CoA and a CoA derivative of unknown structure which would result, perhaps by further modification, in the methoxy group at C-4. It should also be noted that the loading AT for the erythromycin PKS also falls within the methylmalonate class even though the substrate for this AT is propionyl-CoA. Thus, it seems that the methylmalonyl grouping may encompass ATs which recognize a broader range of substrates than previously predicted. As a result, it may be difficult to assign substrate specificity to ATs not assigned to known PKS clusters or to those whose modular positions have not been determined within a known cluster solely on the basis of its homology with other members of the methylmalonyl class.

Domain replacements within the erythromycin PKS have been shown to result in the production of novel bioactive compounds (11, 22, 25). The ethylmalonyl AT domain (AT5) and the methoxy AT domain (AT6) of the niddamycin cluster could conceivably be used to replace the mmAT domains in the erythromycin PKS to generate erythromycin derivatives with novel polyketide backbone structures that would be difficult to produce by chemical methods.

ACKNOWLEDGMENTS

REFERENCES

1. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.
2. Amy, C. M., A. Witkowski, J. Naggert, B. Williams, Z. Randhawa, and S. Smith. 1989. Molecular cloning and sequencing of cDNAs encoding the entire rat fatty acid synthase. Proc. Natl. Acad. Sci. USA 86:3114–3118.
3. Aparicio, J. F., I. Molnar, T. Schwecke, A. Konig, S. F. Haydock, L. E. Khaw, J. Staunton, and P. Leadlay. 1996. Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: analysis of the enzymatic domains in the modular polyketide synthase. Gene 169:9–16.
4. Baltz, R. H. 1980. Genetic recombination in Streptomyces fradiae by protoplast fusion and cell regeneration. Dev. Ind. Microbiol. 21:43–54.
5. Bevitt, D. J., J. Cortes, S. F. Haydock, and P. F. Leadlay. 1992. 6-Deoxyerythronolide B synthase 2 from Saccharopolyspora erythraea. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. Eur. J. Biochem. 204:38–49.
6. Bierman, M., R. Logan, K. O'Brien, E. T. Seno, R. N. Rao, and B. E. Schoner. 1992. Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. Gene 116:43–49.
7. Cortes, J., S. F. Haydock, G. A. Roberts, D. J. Bevitt, and P. F. Leadlay. 1990. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. Nature 348:176–178.
8. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387–395.
9. Donadio, S., M. J. Staver, J. B. McAlpine, S. J. Swanson, and L. Katz. 1991. Modular organization of genes required for complex polyketide biosynthesis. Science 252:675–679.
10. Donadio, S., and L. Katz. 1992. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in Saccharopolyspora erythraea. Gene 111:51–60.
11. Donadio, S., J. B. McAlpine, P. J. Sheldon, M. Jackson, and L. Katz. 1993. An erythromycin analog produced by reprogramming polyketide synthesis. Proc. Natl. Acad. Sci. USA 90:7119–7123.
11a.Fager, E. E., T. J. Oliver, J. F. Prokop, and A. C. Sinclair. November 1965. U.S. patent 3218239.
12. Gandecha, A. R., S. L. Large, and E. Cundliffe. 1997. Analysis of four tylosin biosynthetic genes from the tylLM region of the Streptomyces fradiae genome. Gene 184:197–203.
13. Haydock, S., J. F. Aparicio, I. Molnar, T. Schwecke, A. Konig, A. F. A. Marsden, I. S. Galloway, J. Staunton, and P. F. Leadlay. 1995. Divergent structural motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acylcarrier protein transacylase domains in the modular polyketide synthases. FEMS Lett. 374:246–248.
14. Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. J. Bruton, H. M. Kieser, D. J. Lydiate, C. P. Smith, J. M. Ward, and H. Schempf. 1985. Genetic manipulation of Streptomyces: a laboratory manual. The John Innes Foundation, Norwich, England.
15. Hopwood, D. A., and D. H. Sherman. 1990. Molecular genetics of polyketide and its comparison to fatty acid biosynthesis. Annu. Rev. Genet. 24:37–66.
16. Huber, V. G., K. H. Wallhausser, L. Fries, A. Steigler, and H.-L. Weidenmuller. 1962. Niddamycin, ein neues makrolid-antibioticum. Arzneim. Forsch. 12:1191–1195.
17. Kuhstoss, S., M. Huber, J. R. Turner, J. W. Paschal, and R. N. Rao. 1996. Production of a novel polyketide through the construction of a hybrid polyketide synthase. Gene 183:231–236.
18. Lomovskaya, N., L. Fonstein, X. Ruan, D. Stassi, L. Katz, and C. R. Hutchinson. 1997. Gene disruption and replacement in the rapamycin-producing Streptomyces hygroscopicus strain ATCC 29253. Microbiology 143:875–883.
19. Merson-Davies, L. A., and E. Cundliffe. 1994. Analysis of five tylosin biosynthetic genes from the tylIBA region if the Streptomyces fradiae genome Mol. Microbiol. 13:349–355.
20. Molnar, I., J. F. Aparicio, S. F. Haydock, L. E. Khaw, T. Schwecke, and D. O'Hagan. 1993. Biosynthesis of fatty acid and polyketide metabolites. Nat. Prod. Rep. 10:593–624.
21. Muth, G., B. Nussbaumer, W. Wohlleben, and A. Puhler. 1989. A vector system with temperature-sensitive replication for gene disruption and mutational cloning in streptomyces. Mol. Gen. Genet. 219:341–348.
22. Oliynyk, M., M. J. B. Brown, J. Cortes, J. Staunton, and P. F. Leadlay. 1996. A hybrid modular polyketide synthase obtained by domain swapping. Chem. Biol. 3:833–839.
23. Omura, S., A. Nakagawa, H. Takeshima, K. Atsumi, J. Miyazawa, F. Piriou, and G. Lukacs. 1975. Biosynthetic studies using ^{13}C enriched precursors on the 16-membered macrolide antibiotic leucomycin A$_3$. J. Am. Chem. Soc. 97:6600–6602.
24. Omura, S., K. Tsuzuki, A. Nakagawa, and G. Lukacs. 1983. Biosynthetic origin of carbons 3 and 4 of leucomycin aglycone. J. Antibiot. 36:611–613.
25. Ruan, X., A. Pereda, D. L. Stassi, D. Zeidner, R. G. Summers, M. Jackson, A. Shivakumar, S. Kakavas, M. J. Staver, S. Donadio, and L. Katz. 1997. Acyltransferase domain substitutions in the erythromycin polyketide synthase yield novel erythromycin derivatives. J. Bacteriol. 179:6416–6425.
26. Ruan, X., D. Stassi, S. Lax, and L. Katz. A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC 29253, a rapamycin-producing strain. Gene, in press.
27. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. Schwecke, T., J. F. Aparicio, I. Molnar, A. Konig, L. E. Khaw, S. F. Haydock, M. Oliynyk, P. Caffrey, J. Cortes, J. B. Lester, G. A. Bohm, J. Staunton, and P. F. Leadlay. 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. Proc. Natl. Acad. Sci. USA 92:7839–7843.
29. Scrutton, N. S., A. Berry, and R. N. Perham. 1990. Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. Nature 343:38–42.
30. Strohl, W. R. 1992. Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. Nucleic Acids Res. 20:961–974.
31. Tuan, J. S., J. M. Weber, M. J. Staver, J. O. Leung, S. Donadio, and L. Katz. 1990. Cloning of genes involved in erythromycin biosynthesis from Saccharopolyspora erythraea using a novel actinomycete-Escherichia coli cosmid. Gene 90:21–29.
32. Witkowski, A., V. S. Rangan, Z. I. Randhawa, C. M. Amy, and S. Smith. 1991. Structural organization of the multifunctional animal fatty-acid synthase. Eur. J. Biochem. 198:571–579.
33. Wright, F., and M. J. Bibb. 1992. Codon usage in the G+C-rich Streptomyces genome. Gene 113:55–65.
34. Yamamoto, H., K. H. Maurer, and C. R. Hutchinson. 1986. Transformation of Streptomyces erythraeus. J. Antibiot. 39:1304–1313.
35. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.

Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering

D. L. Stassi[†‡], S. J. Kakavas[†], K. A. Reynolds[§], G. Gunawardana[†], S. Swanson[†], D. Zeidner[†], M. Jackson[†], H. Liu[§], A. Buko[†], and L. Katz[†]

[†]Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, IL 60064; and [§]Institute of Structural Biology and Drug Discovery, Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23219

ABSTRACT   A previously unknown chemical structure, 6-desmethyl-6-ethylerythromycin A (6-ethylErA), was produced through directed genetic manipulation of the erythromycin (Er)-producing organism Saccharopolyspora erythraea. In an attempt to replace the methyl side chain at the C-6 position of the Er polyketide backbone with an ethyl moiety, the methylmalonate-specific acyltransferase (AT) domain of the Er polyketide synthase was replaced with an ethylmalonate-specific AT domain from the polyketide synthase involved in the synthesis of the 16-member macrolide niddamycin. The genetically altered strain was found to produce ErA, however, and not the ethyl-substituted derivative. When the strain was provided with precursors of ethylmalonate, a small quantity of a macrolide with the mass of 6-ethylErA was produced in addition to ErA. Because substrate for the heterologous AT seemed to be limiting, crotonyl-CoA reductase, a primary metabolic enzyme involved in butyryl-CoA production in streptomycetes, was expressed in the strain. The primary macrolide produced by the reengineered strain was 6-ethylErA.

Erythromycin (Er) is a broad-spectrum macrolide antibiotic produced by Saccharopolyspora erythraea. The backbone of the molecule is a 14-member macrocyclic ring (Fig. 1) that is produced through the sequential condensation of one molecule of propionyl-CoA and six molecules of methylmalonyl-CoA by a modular polyketide synthase (PKS). This enzyme complex comprises three large multifunctional polypeptides, DEBS1, DEBS2, and DEBS3, each of which contains two modules, and, in the case of DEBS1, a loading domain to initiate synthesis of the polyketide chain (1, 2). Each module contains the enzymatic activities necessary for one condensation and subsequent reduction of the extender to the growing chain. Selection of the appropriate extender unit is accomplished by the acyltransferase (AT) domain present in each module (3).

Recently, genetic engineering has joined synthetic chemistry in the production of novel macrolide structures. Hybrid PKSs have been constructed through the replacement of AT domains with those that specify different starter or extender units (3–6). These manipulations have resulted in the production of Er derivatives lacking methyl groups at C-10 and C-12 (5) and those incorporating the branched chain starter units of avermectin biosynthesis (6). To date, however, there have been no manipulations that allow the methyl side chains of Er to be replaced by ethyl groups. This type of modification would be useful in expanding the structural diversity of hybrid polyketides produced by combinatorial biosynthesis.

Niddamycin (Nd) is a 16-member macrolide with an ethyl side chain at C-6 (Fig. 1). The module 5 AT, which is believed to be responsible for incorporation of the ethyl side chain into the polyketide backbone, was identified by sequence analysis of the Nd PKS genes (7). In this paper we describe the construction of a hybrid PKS through the replacement of a methylmalonyl-specific AT of the Er PKS with that of the ethylmalonyl-specific AT of the Nd PKS. This substitution alone, however, was not sufficient to produce an Er derivative with an ethyl side chain; additional manipulations of the carboxylic acid precursor pools were necessary to generate the desired compound.

MATERIALS AND METHODS

Strains, Plasmids, and Media. The wild-type Er producer is S. erythraea ER720 (8). Plasmid pWHM3 (9) is an Escherichia coli–Streptomyces shuttle vector that is maintained in S. erythraea only when it contains heterologous DNA for chromosomal integration. Plasmid pDPE81 is a derivative of pKAS37 (10) in which a 1.7-kb BglII fragment containing the hygromycin resistance marker was inserted into the BglII site of pKAS37. E. coli DH5α (Life Technologies, Gaithersburg, MD) was the host used for plasmid construction and isolation. Liquid cultures of S. erythraea strains were grown in SGGP (11) for production of protoplasts and SCM (5) for metabolite or enzyme analysis. Plate cultures of S. erythraea strains were grown on R3M medium (5). Thiostrepton (Ts)-resistant strains of S. erythraea were grown in 25 μg/ml and 10 μg/ml Ts for plate and liquid cultures, respectively.

Plasmid pEAT4 (Fig. 2A) was constructed as follows. Cosmid pAIBX85, a pWHM3 derivative containing DNA from modules 3 and 4 of the Er PKS (corresponding to nucleotides 979-9349; GenBank accession no. M63677), was used to clone DNA flanking eryAT4. The 5' flanking region was isolated by digesting pAIBX85 with MscI and BstEII (nucleotides 4247–6033), treating with the Klenow fragment of DNA polymerase, and ligating the fragment into the SmaI site of pUC19 to generate pUC/5' flank. An AvrII site was engineered 13 bp downstream of the BstEII site by PCR amplification of a 306-bp region of DNA from the PmlI site (nucleotide 5739) to 12 bp 3' of the BstEII site (nucleotide 6045). The engineered AvrII site does not change the Pro-Arg residues encoded by this region (Fig. 2B). A BamHI site was also included on the PCR primer just downstream of the AvrII site. The resulting fragment was digested with PmlI and BamHI and cloned into the PmlI/BamHI site of pUC/5' flank, replacing the native se-

FIG. 1.   Structures of Er, Nd, and 6-ethylErA. Shadowed letters represent relevant side groups.

Erythromycin A   Niddamycin   6-Desmethyl-6-ethylerythromycin A

quence and resulting in vector pUC/5' flank/*Avr*II. The 3' flanking DNA was isolated by digesting pAIXB85 with *Pml*I and *Msc*I (nucleotides 6999 and 8977), treating with the Klenow enzyme, and ligating the fragment into the *Sma*I site of pUC19, generating vector pUC/3' flank. nidAT5 was isolated by PCR amplification of cosmid 13f5 (7), which contains DNA spanning module 5 of the Nd PKS cluster. The 5' oligonucleotide was designed to create an *Avr*II site 12 nucleotides upstream of the beginning of the nidAT5 domain (nucleotide 25290; GenBank accession no. AF016585) and to place the amino acid sequence Pro-Arg-Lys-Pro in front of nidAT5 to correspond to the sequence that is found upstream of the eryAT4 domain. The 3' oligonucleotide generated an *Fse*I site at the end of the nidAT5 domain (nucleotide 26284, GenBank accession no. AF016585), resulting in a conservative Val to Ala change. A *Bam*HI site was also incorporated after the *Fse*I site. The fragment was digested with *Avr*II and *Bam*HI and ligated into *Avr*II/*Bam*HI-digested pUC/5' flank/*Avr*II, creating vector pUC/5' flank/nidAT5. The 3' flanking DNA was then cloned onto the 3' end of nidAT5 by first digesting pUC/3' flank with *Fse*I and *Bam*HI, gel purifying the 1920-bp

fragment, and ligating it into *Fse*I/*Bam*HI-digested pUC/5' flank/nidAT5, creating vector pUC/nidAT5/C6-flank. To generate pEAT4, the nidAT5/flanking DNA cassette was isolated from pUC/nidAT5/C6-flank by digestion with *Eco*RI and *Hin*dIII and then ligated to *Eco*RI/*Hin*dIII-digested pWHM3.

Plasmid pDPE-ccr (Fig. 3) was constructed as follows. The *Streptomyces collinus* crotonyl-CoA reductase (Ccr) gene (*ccr*) was subcloned from plasmid pZYB3 (12) by digestion with *Xba*I and *Bam*HI, which releases *ccr* and the upstream T7 ribosomal binding site. This fragment was treated with the Klenow enzyme and ligated into the polylinker of pDPE81 that had been digested with *Eco*RI and treated with the Klenow enzyme. The polylinker of pDPE81 is in the center of a 10-kb fragment of *S. erythraea* chromosomal DNA, which directs integration into the chromosome at a site that does not seem to affect Er production. Plasmid pDPE-ccr was designed so that the *ccr* gene expressed from the *ermE** promoter (13) and the Ts-resistance marker are left behind in the chromosome following a double-crossover event.

Genetic Manipulations. Standard molecular biology techniques were performed as described (14). Enzymes and reagents were purchased from Life Technologies. Protoplast transformation and marker replacement in *S. erythraea* were performed as described (5). One microgram of plasmid DNA was routinely used for protoplast transformations. Putative transformants were grown in SGGP medium containing 10 μg/ml Ts to confirm resistance. For chromosomal eviction of pEAT4, transformants were passaged twice in SGGP medium without Ts, cells were plated for spores, and individual colonies arising from spores were screened for Ts sensitivity.

A

4247   6045   nidAT5   7045   8977
*Eco*RI -[████████████]*Avr*II ——————*Fse*I-[████████████]-*Hin*dIII

B

N-terminus

eryAT4 wild type	APVTSAPRKP	VLVFPGQGAQ
nidAT5 wild type	TGTAGTADRV	VFVFPGQGSQ
eryAT4 flank/nidAT5	APVTSAPRKP	VFVFPGQGSQ

↓

*Avr*II site

C-terminus

eryAT4 wild type	ADLRPAVAGG	RPAELPTYPE
nidAT5 wild type	DPEAVFGPGA	RPVELPTYPF
eryAT4 flank/nidAT5	DPEAVFGPGA	RPAELPTYPF

↓

*Fse*I site

FIG. 2.   (*A*) Diagrammatic representation of insert in vector pEAT4. DNA encoding the AT from module 5 of the Nd PKS cluster (nidAT5) was cloned between DNA flanking the 5' and 3' boundaries of the AT coding region in module 4 of the Er PKS cluster. The numbered boxes correspond to *eryA* sequence coordinates from GenBank accession no. M63676. (*B*) Amino acid comparisons in the junction regions of eryAT4, nidAT5, and the eryAT4 flank/nidAT5 construct. The amino acids encoding the *Avr*II and *Fse*I sites are indicated.

FIG. 3.   Plasmid pDPE-ccr. The thick black lines represent *S. erythraea* DNA, which allows for integration of the plasmid by homologous recombination into the *S. erythraea* chromosome. Also indicated are genes for ampicillin (amp) and Ts resistance.

Biochemistry: Stassi *et al.*

Proc. Natl. Acad. Sci. USA 95 (1998)   7307

Electroporation was performed as described (15) by using 1 μg of pDPE-ccr DNA. Because the plasmid is unstable when integrated into the *S. erythraea* chromosome, two consecutive platings on Ts-containing R3M plates results in resolution and eviction of plasmid sequences. PCR and Southern hybridization were as described (7).

Ccr Assay. Seed cultures of *S. erythraea* grown in SCM and *S. collinus* grown in medium A (16) were diluted 5-fold into SCM and medium A, respectively, and grown for 48 h at 30°C. Cells were harvested, opened with a French pressure cell, centrifuged to obtain cell-free extracts, and assayed spectrophotometrically for Ccr activity as described (16). One unit of Ccr activity is defined as 1 μmol of NADPH oxidized in 1 min.

Isolation and Identification of Metabolites. Small-scale (milliliters) isolation of metabolites from *S. erythraea*, TLC, and MS analysis were performed as described (5). Large-scale isolation for compound identification was as follows. Fermentations were conducted in 42-liter LH fermentation series 2000 stainless steel vessels (LH Fermentation, Maidenhead, UK). Each fermentor was charged with 30 liters of a medium consisting of 1.5% soluble starch, 2.2% soybean flour, 0.2% CaCO₃, 0.15% brewer's yeast (Wind Gap Farms, Baconton, GA), 0.102% MgSO₄·7H₂O, 0.0027% FeSO₄·7H₂O, and 1% soybean oil. Sterilization was at 121°C and 15 psi (1 psi = 6.89 kPa). A two-step seed protocol was used. Vegetative growth from a frozen stock of *S. erythraea* strain EAT4-ccr was inoculated at 1.5% into a 500-ml Erlenmeyer flask containing 100 ml of the following medium: 1.5% glucose monohydrate/1% soluble starch/1% Soytone (Difco)/0.9% yeast extract (Difco)/0.21% Mops/0.001% Ts. For the second step, a 2-liter Erlenmeyer flask containing 600 ml of the same medium was inoculated at 5% with the first-step growth. Both steps were incubated at 28°C on a rotary shaker at 225 rpm for 48 h. The fermentor was inoculated at 5% with the second-step seed growth. Fermentor temperature was controlled at 32°C, agitation was 250 rpm, aeration was 1 vol·vol⁻¹·min⁻¹, and head pressure was 6 psi. Silicone antifoam was added at 0.01% initially and was available on demand. Harvest was at 108 h.

The fermentation broth (28 liters) was filtered, the pH was adjusted to 9 with NH₄OH, and the broth was extracted with CH₂Cl₂ (twice with 14 liters each). The combined extracts were concentrated, and the residue was partitioned between the two phases of the heptane/methanol/0.02 M K₂HPO₄ system (pH 6; 1:1:1). The aqueous phase was evaporated, and the residue was dissolved in 0.05 M aqueous potassium phosphate buffer (100 ml), adjusted to pH 9 with NH₄OH, and extracted twice with equal volumes of CH₂Cl₂. The CH₂Cl₂-soluble fraction (1.65 g) was separated by droplet countercurrent chromatography (100 vertical columns; 0.4 × 24 cm) by using the upper layer from the hexane/ethyl acetate/0.02 M K₂HPO₄ (pH 8; 1:1:1) system as th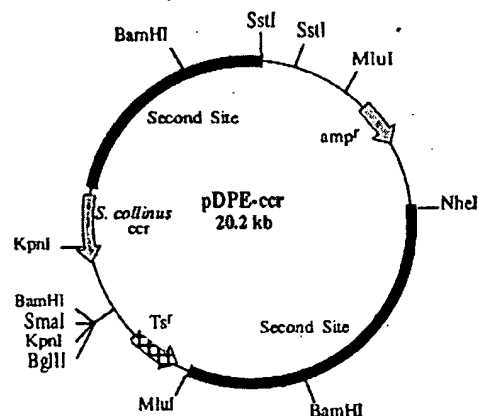e mobile phase. The fractions were analyzed by bioassay against *Staphylococcus aureus* and by ¹H NMR. The bioactive fractions that were shown to contain macrolides by NMR were pooled, dried (0.4 g), and chromatographed on a Sanki Engineering (Kyoto) HPLC centrifugal partition chromatograph by using the hexane/ethyl acetate/0.02 M K₂HPO₄ (pH 6; 1:1:1) system. The active fraction was further purified by HPLC on C₁₈-derivatized silica by using an acetonitrile/methanol/0.01 M (CH₃)₄NOH/0.05 M KH₂PO₄ (pH 6; 73:10:59:59) system. Two-dimensional NMR experiments, including double-quantum correlation, heteronuclear multiple quantum correlation, and heteronuclear multiple-bond correlation, performed in C²HCl₃, were used for structure elucidation and the complete assignment of the ¹H and ¹³C NMR signals.

In Vitro Antibacterial Activity. Antibacterial activity was determined by the broth microdilution method (National Committee for Clinical Laboratory Standards (Villanova, PA) M7-A4, 1997), except that Brain Heart Infusion (Difco) broth was used as the test medium. Assays were incubated overnight at 35°C. Minimal inhibitory concentrations were defined as the lowest drug concentration (μg/ml) inhibiting bacterial growth.

RESULTS

Construction of *S. erythraea* EAT4. Plasmid pEAT4 was constructed to replace DNA encoding the AT domain of module 4 in the Er PKS (eryAT4) with DNA encoding the AT domain from module 5 of the Nd PKS (nidAT5) (7). *S. erythraea* ER720 protoplasts were transformed with pEAT4 DNA, and 10 transformants were obtained. Genomic DNA was extracted from one of the transformants for Southern analysis, in which probing with pWHM3 DNA confirmed the integration of pEAT4 at the appropriate location in the chromosome (data not shown). Nonselective growth to allow plasmid eviction through a double-crossover event yielded 96 colonies, of which 9 were found to be Ts sensitive. Southern analysis of the 9 clones showed that 3 had nidAT5 DNA sequences in place of eryAT4 chromosomal sequences, and the remaining 6 had segregated to wild type (data not shown).

Characterization of *S. erythraea* EAT4. To analyze the Er derivatives produced by *S. erythraea* EAT4, the three isolates and four of the wild-type segregants were grown in SCM medium for 4 d. The four wild-type segregants produced spots indistinguishable in color and R_f from the ErA standard. The three strains in which nidAT5 replaced the AT of module 4 of the Er PKS produced spots similar in R_f and color to the wild-type segregants, but the spots were much less intense (data not shown).

To determine the mass of the compound produced by *S. erythraea* EAT4, the supernatant of a 50-ml SCM culture of one of the isolates was extracted with ethyl acetate. The extract was subjected to TLC, but only the edges of the plate were sprayed with anisaldehyde to locate the region of interest. A 1-cm band of resin was scraped from the unsprayed portion of the plate at the R_f of ErA. The resin was extracted twice with 500 μl of an ethyl acetate/methanol (2:1) solution, and the organic phase was dried and then analyzed by electrospray ionization MS. Surprisingly, the mass of the protonated molecular ion of the compound was observed at m/z 734, which corresponds to the mass of ErA, not an ethyl-substituted ErA derivative.

Feeding of Ethylmalonyl-CoA Precursors to *S. erythraea* EAT4. One hypothesis for the failure of *S. erythraea* EAT4 to produce an ethyl-substituted derivative is that ethylmalonyl-CoA, the substrate for the nidAT5, is lacking. To test this hypothesis, cells were grown for 4 d in SCM medium containing precursor compounds, and ethyl acetate extracts of the cultures were analyzed. When either 50 mM butanol or 50 mM butyrate was added to the medium, a second spot running slightly faster than the R_f of ErA was seen. Butyrate-fed cultures (Fig. 4*A*) produced about a 1:1 ratio of the two spots, whereas the butanol-fed culture produced more of the ErA-like spot (data not shown). Addition of 10 mM ethylmalonate failed to produce significant amounts of either of the two spots described above. However, addition of 10 mM diethylethylmalonate was found to yield much more of the faster migrating compound than that seen with the other precursor compounds tested and very little compound migrating at the R_f of ErA (Fig. 4*A*). Cells grown in unsupplemented SCM medium produced only material migrating like ErA.

To characterize the newly synthesized compound, extracts of the butyrate- and the diethylethylmalonate-fed cultures were subjected to TLC, the region of interest was scraped from the plate and reextracted for electrospray ionization MS. The results showed that the butyrate-fed culture of *S. erythraea* EAT4 produced approximately equal amounts of compounds with protonated molecular ions at m/z 734 and 748. The 748 species is consistent with an additional methylene group on ErA, e.g., with an ethyl group replacing a methyl group on the

A.   **B**

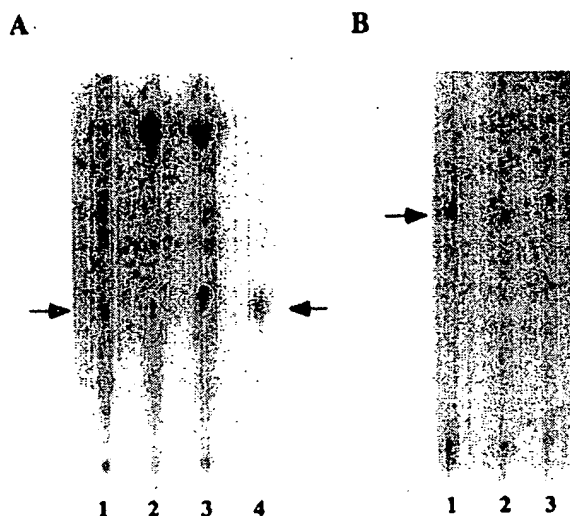**1 2 3 4   1 2 3**

FIG. 4.   TLC analysis of *S. erythraea* EAT4 and *S. erythraea* EAT4-ccr. (*A*) Ethyl acetate extracts of *S. erythraea* EAT4 cultures grown under various conditions. Lanes: 1, SCM medium; 2, SCM + 50 mM butyrate; 3, SCM + 10 mM diethylethylmalonate; 4, ErA standard (5 μg). The arrows indicate the position of ErA. (*B*) Ethyl acetate extracts of *S. erythraea* strains EAT4 (lane 2) and EAT4-ccr (lane 3) grown in SCM medium, along with 5.0 μg of ErA standard (lane 1). The arrow indicates the position of ErA.

macrolide ring. The diethylethylmalonate sample contained primarily a compound with a mass of 748, with only a trace of ErA (734) present.

Genetic Manipulation of Ethylmalonyl-CoA Levels in *S. erythraea* EAT4. Successful production of an Er derivative with a mass consistent with the addition of an ethyl side chain by butyrate- and diethylmalonate-fed *S. erythraea* EAT4 suggested that the levels of ethylmalonyl-CoA, the likely substrate of the nidAT5, must be limiting. In streptomycetes, Ccr catalyzes the last step in the reductive biosynthesis of butyryl-CoA from two molecules of acetyl-CoA (12). Because butyryl-CoA can then be carboxylated to form ethylmalonyl-CoA, the possibility that Ccr could be used to increase the levels of the ethyl-substituted Er derivative was investigated. The *ccr* gene of *S. collinus* was expressed from the strong *ermE** promoter at a site unlinked to the Er biosynthetic cluster in the *S. erythraea* EAT4 chromosome. Plasmid pDPE-ccr was electroporated into *S. erythraea* EAT4. About 40 Ts-resistant colonies were obtained, of which 7 were confirmed to be stable transformants. Genomic DNA was isolated from 2 of the transformants for Southern analysis. Probing with the *S. collinus* ccr gene confirmed that both isolates carried the gene at the expected location in the chromosome (data not shown). The strain was named *S. erythraea* EAT4-ccr.

To assess expression levels of the *ccr* gene, Ccr activity of *S. erythraea* EAT4-ccr was compared with that of *S. erythraea* EAT4 (parental strain) and wild-type *S. collinus*. We found (Table 1) that, although there was no detectable activity in the *S. erythraea* strain without the *ccr* gene, *S. erythraea* EAT4-ccr

Table 1.   Ccr activity in *S. collinus* and engineered *S. erythraea* strains

Sample	Enzyme activity, milliunits/ml	Protein concentration, mg/ml	Specific activity, milliunits/mg
S. erythraea EAT4	<0.5 ± 0.0	3.0	<0.2 ± 0.0
S. erythraea EAT4-ccr	46.9 ± 2.0	3.0	15.6 ± 0.7
S. collinus wild type	5.5 ± 1.0	7.0	0.8 ± 0.1

had about 20 times the relative activity of the wild-type *S. collinus* strain from which the gene originated.

To determine the effect of the *ccr* gene on production of an ethyl-substituted Er derivative, ethyl acetate extracts of supernatants of 4-d SCM-grown cultures of the EAT4 strains with and without the *ccr* gene were examined (Fig. 4*B*). TLC analysis of 10 ml of extracted cells demonstrated that EAT4 without the *ccr* gene produced a spot that migrated with ErA, whereas EAT4-ccr produced a spot that migrated faster than the ErA spot, with no ErA-like material visible.

Extracts were also analyzed by electrospray ionization MS. The EAT4 strain produced a compound with a protonated molecular ion at m/z 734, indicating ErA production. The EAT4-ccr strain produced a compound with a mass of 748, which is consistent with the production of the 6-ethylErA. No compound with a mass of 734 (ErA) was detected in extracts of this strain.

Structural Analysis and Biological Activity of Compound Produced by *S. erythraea* EAT4-ccr. The structure of 6-ethyl-ErA was confirmed by spectroscopic methods. The high resolution fast atom bombardment MS analysis of the sample gave an M + H$^+$ ion at m/z 748.4846 [calculated for $C_{38}H_{70}NO_{13}$ m/z = 748.4842 (Δ 0.5 ppm)], indicating that this molecule has an additional methylene function compared with ErA. The presence of an ethyl moiety in the molecule at the C-6 position and its relative stereochemistry were confirmed

Table 2.   NMR data of 6-ethylErA

Carbon no.	^{13}C shift (δ), ppm	^{1}H shift (δ), ppm
1	175.5	
2	44.8	2.93
3	80.7	3.79
4	37.6	1.93
5	81.6	3.86
6	76.7	
7	38.4	1.72, 1.93
8	44.7	2.70
9	222.4	
10	37.7	3.01
11	68.9	3.82
12	74.4	
13	76.7	5.06
14	21.2	1.47, 1.91
15	10.7	0.84
2-CH₃	15.9	1.29
4-CH₃	9.7	1.14
6-CH₂-CH₃	27.6	1.70, 2.02
6-CH₂-CH₃	7.7	0.98
8-CH₃	18.6	1.17
10-CH₃	12.3	3.01
12-CH₃	68.9	3.82
1'	102.5	4.48
2'	72.5	3.21
3'	65.3	2.44
4'	28.3	1.22, 1.63
5'	68.7	3.52
6'	21.6	1.22
3'-N(CH₃)₂	40.7	2.29
1″	96.2	4.92
2″	35.3	2.34, 1.66
3″	72.4	
4″	77.7	3.02
5″	66.3	3.98
6″	21.6	1.29
3″-OCH₃	49.7	3.33
3″-CH₃	21.4	1.26

Numbering of carbon atoms of Er is as described (17).

by two-dimensional NMR. The stereospecificity of the ethyl moiety was found to be the same as that of the methyl group present at C-6 of ErA (Fig. 1). The ^{1}H and ^{13}C NMR assignments are shown in Table 2.

Antibacterial activity of 6-ethylErA was compared with that of ErA against a panel of *S. aureus, Staphylococcus epidermidis, Streptococcus pyogenes,* and *Enterococcus faecium* strains from the Abbott culture collection. Although 6-ethylErA does retain biological activity, it was found to be 15 to 60 times less potent than ErA against these strains. Representative minimal inhibitory concentrations (μg/ml) for 6-ethylErA and ErA, respectively, were as follows: *S. aureus* (4 and 0.06), *S. epidermidis* (4 and 0.12), *E. faecium* (4 and 0.12), and *S. pyogenes* (1 and 0.06).

DISCUSSION

In this work we have shown that an ethylmalonate AT can be substituted for a methylmalonate AT to produce a hybrid PKS that synthesizes the predicted macrolide product, 6-ethylErA. This suggests that Er analogs containing ethyl substitutions at positions other than C-6 or that ethyl analogs of other macrolides, such as rifamycin or rapamycin, can be prepared in a similar fashion by using the ethylmalonate-specific AT from the Nd PKS.

One key feature for the successful generation of 6-ethylErA was availability of ethylmalonyl-CoA as substrate for the ethylmalonate-specific AT of the Nd PKS in *S. erythraea*. This AT clearly demonstrates a relaxed specificity by reacting with both ethylmalonyl-CoA and methylmalonyl-CoA. In *S. erythraea* EAT4, the desired product could be obtained only by boosting the levels of ethylmalonyl-CoA by either precursor feeding or expression of Ccr, an enzyme involved in butyryl-CoA biosynthesis. These results suggest that ethylmalonyl-CoA derived from butyryl-CoA is not available at sufficient levels to compete with methylmalonyl-CoA for the ethylmalonate-specific AT. Ccr activity is not detectable in *S. erythraea* and attempts to detect a *ccr* gene in this organism by hybridization with the *S. collinus ccr* gene have failed (results not shown). In contrast, the *ccr* gene appears to be present in most streptomycetes (16), and in certain cases it appears to be clustered with antibiotic biosynthetic gene clusters. For example, a *ccr* homolog has been located in the biosynthetic cluster for tylosin (18), another 16-member macrolide with an ethyl side chain. Southern analysis of cosmid clones containing Nd biosynthetic genes also indicates that a *ccr* homolog may reside close to the Nd PKS (unpublished results).

It was demonstrated previously (5) that production of desmethyl Er derivatives by replacement of methylmalonate-specific ATs of the Er PKS with malonate-specific ATs from *Streptomyces hygroscopicus* and *Streptomyces venezuelae* was successful only in Er modules 1 and 2 (corresponding to C-12 and C-10 of Er, respectively) and did not give rise to a detectable polyketide when placed in module 4. In contrast, we successfully produced ethyl-substituted Er by replacement of the methylmalonate-specific AT of Er module 4 with the ethylmalonate-specific AT from the Nd cluster. In fact, in *S. erythraea* expressing the *S. collinus ccr* gene, replacements in modules 1 through 4 with the ethylmalonate-specific AT produced compounds with electrospray ionization mass spectra consistent with production of ethyl substitutions for methyl groups at C-12, C-10, C-8, and C-6, respectively. However, those in modules 5 and 6 did not produce compounds of mass 748 (data not shown). It is not known whether failure to produce the predicted compounds was caused by physical distortion of the PKS or its mRNA, inability of the PKS to process certain altered growing chains, or some structural instability of the macrolide itself because of the introduced change. Thus, even though genetic information may be present for the production of novel compounds, it is still not possible to predict which substitutions will yield detectable levels of product.

It has also been demonstrated that the malonyl-specific ATs can be distinguished from methylmalonyl-specific ATs through sequence alignments (3). By using this strategy, Ruan *et al.* (5) predicted that an AT from an unidentified PKS of *S. hygroscopicus* was a malonyl-specific AT, and they used it successfully to produce desmethyl Er derivatives. The Nd ethyl-specific AT was found to cluster with methylmalonyl-specific ATs (7) in similar AT alignments. This similarity may explain why methylmalonate was used as a substrate by nidAT5 in *S. erythraea* EAT4 to produce ErA when ethylmalonyl-CoA was not available.

Finally, structural determination by NMR not only confirmed the presence of the ethyl side chain on the Er derivative produced by *S. erythraea* EAT4 but also showed that the absolute configuration at this chiral center is the same as that in ErA. In Er biosynthesis, epimerization is required at C-6 because the PKS uses (2*S*)-methylmalonyl-CoA for chain extension (19). It is unclear whether the cognate epimerization occurs when ethylmalonate is used in the synthesis of 6-ethylErA because the stereochemistry of the ethylmalonate incorporated at C-6 is not known.

1. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. (1990) *Nature (London)* **348**, 176–178.
2. Donadio, S. & Katz, L. (1992) *Gene* **111**, 51–60.
3. Oliynyk, M., Brown, M. J. B., Cortes, J., Staunton, J. & Leadlay, P. F. (1996) *Chem. Biol.* **3**, 833–839.
4. Kuhstoss, S., Huber, M., Turner, J. R., Paschal, J. W. & Rao, R. N. (1996) *Gene* **183**, 231–236.
5. Ruan, X., Pereda, A., Stassi, D., Zeidner, D., Summers, R., Jackson, M., Shivakumar, A, Kakavas, S. & Katz, L. (1997) *J. Bacteriol.* **179**, 6416–6425.
6. Marsden, A. F. A., Wilkinson, B., Cortes, J., Dunster, N. J., Staunton, J. & Leadlay, P. F. (1998) *Science* **279**, 199–202.
7. Kakavas, S. J., Katz, L. & Stassi, D. (1997) *J. Bacteriol.* **179**, 7515–7522.
8. DeWitt, J. P. (1985) *J. Bacteriol.* **164**, 969–971.
9. Vara J. A., Lewandowska-Skarbek, M., Wang, Y.-G., Donadio, S. & Hutchinson, C. R. (1989) *J. Bacteriol.* **171**, 5872–5881.
10. Maine, G., Post, D., Satter, M. & Stassi, D. (1997) International patent application WO 97/06266 (2/20/97).
11. Yamamoto, H., Maurer, K. H. & Hutchinson, C. R. (1986) *J. Antibiot.* **39**, 1304–1313.
12. Wallace, K. K., Han, L., Lobo, S. & Renolds, K. K. (1997) *J. Bacteriol.* **179**, 3884–3891.
13. Bibb, M. J., White, J., Ward, J. M. & Janssen, G. R. (1994) *Mol. Microbiol.* **14**, 533–545.
14. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
15. Fitzgerald, N. B., English, R. S., Lampel, J. S. & Vanden Boom, T. (1998) *Appl. Environ. Microbiol.*, **64**, 1580–1583.
16. Wallace, K. K., Bao, Z.-Y., Dai, H. Digate, R., Schuler, G., Speedie, M. & Reynolds, K. (1995) *Eur. J. Biochem.* **233**, 954–962.
17. Larty, P. A. & Faghih, R. (1993) in *Recent Progress in the Chemical Synthesis of Antibiotics,* eds. Lukacs, G & Ohno, M. (Springer, New York), p. 121.
18. Gandecha, A. R., Large, S. L. & Cundliffe. E. (1997) *Gene* **184**, 197–203.
19. Weissman, K. J., Timoney, M., Bycroft, M., Grice, P., Hanefeld, U., Staunton, J. & Leadlay, P. F. (1997) *Biochemistry* **36**, 13849–13855.

51

Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*

(FAS; fatty acids; macrolide antibiotic; sequence alignments; *Streptomyces*)

Stefano Donadio and Leonard Katz

Corporate Molecular Biology, Abbott Laboratories, Abbott Park, IL 60064 (U.S.A.)

SUMMARY

Localization of the enzymatic domains in the three multifunctional polypeptides from *Saccharopolyspora erythraea* involved in the formation of the polyketide portion of the macrolide antibiotic erythromycin was determined by computer-assisted analysis. Comparison of the six synthase units (SU) from the *eryA* genes with each other and with mono- and multifunctional fatty acid and polyketide synthases established the extent of each β-ketoacyl acyl-carrier protein (ACP) synthase, acyltransferase, β-ketoreductase, ACP, and thioesterase domain. The extent of the enoyl reductase (ER) domain was established by detecting similarity to other sequences in the database. A segment containing the putative dehydratase (DH) domain in EryAII, with a potential active-site histidine residue, was also found. The finding of conservation of a portion of the DH-ER interdomain region in the other five SU, which lack these two functions, suggests a possible evolutionary path for the generation of the six SU.

INTRODUCTION

Erythromycin, a macrolide antibiotic produced by *S. erythraea*, is composed of the polyketide-derived 14-membered macrolactone ring, 6dEB, to which are attached two deoxysugars, cladinose and desosamine. Synthesis of 6dEB

Correspondence to: Dr. L. Katz, Abbott Laboratories, D-93D, One Abbott Park Rd., Abbott Park, IL 60064 (U.S.A.)
Tel. (708)937-4132; Fax (708)938-6046.

Abbreviations: aa, amino acid(s); ACP, acyl-carrier protein; ACP-S, ACP of SU1; AT, acyltransferase; CoA, coenzyme A; 6dEB, 6-deoxyerythronolide B; DH, dehydratase; *dnaB*, gene encoding helicase; ER, enoyl reductase; *ery*, erythromycin biosynthesis gene; *eryA*, gene encoding 6dEB synthase; FAS, fatty acid synthase; FAS1, *S. cerevisiae* FAS β-chain; KR, β-ketoreductase; KS, β-ketoacyl ACP synthase; 6MSAS, 6-methylsalicylic acid synthase; ORF, open reading frame; PKS, polyketide synthase; *S.*, *Saccharopolyspora*; SU, synthase unit(s); TE, thioesterase; URF, unidentified ORF.

involves six elongation cycles that resemble the steps in fatty acid synthesis. It has recently been shown that 6dEB synthesis requires three adjacent *eryA* genes encoding large multifunctional polypeptides and that the *eryA* cluster consists of six modules (repeated motifs), each encoding a different SU specific for one of the elongation steps (Cortes et al., 1990; Donadio et al., 1991). We proposed that the genetic organization of *eryA* and the steps in the biochemical pathway of 6dEB are colinear (Donadio et al., 1991). A scheme of the enzymatic activities leading to 6dEB is shown in Fig. 1.

In previous work, the FAS-like activities ACP, AT, KR and KS in the multifunctional *eryA*-encoded polypeptides were identified from the presence of 'signature sequences' found at the active site of the functional domains (Cortes et al., 1990; Donadio et al., 1991). However, signature sequences have not been assigned previously to the DH and ER functions, both of which have been poorly characterized biochemically. Multifunctional FAS systems are or-

Fig. 1. Scheme of 6-deoxyerythronolide B synthesis (Donadio et al., 1991). The top portion shows the three *eryA*-encoded polypeptides containing SU 1 through 6. Enzymatic activities belonging to the first and to the second SU for each polypeptide are represented by empty and shaded circles, respectively. The bottom portion illustrates the role of each SU in the synthesis of 6dEB, where the C_2 units in the ring introduced by odd- and even-numbered SU are represented by dashed and continuous lines, respectively.

ganized in discrete functional domains which can be resolved upon limited proteolysis (Wakil, 1989). The *eryA*-determined SU can be assumed to consist of linear domains on the basis of their similar organization to the animal FAS systems (Cortes et al., 1990; Donadio et al., 1991). Here, by comparing the 6dEB PKS domains with each other and with other multi- and monofunctional FAS and PKS systems, we subdivide each *eryA*-encoded polypeptide into its constituent domains and propose a location for the DH and ER domains in EryAII. Conclusions similar to those reported here have been independently reached in the laboratory of P.F. Leadlay (personal communication) for the domain organization of EryAII and EryAIII and by Witkowski et al. (1991a) for the domain organization of rat FAS.

RESULTS AND DISCUSSION

(a) Extent of KS, AT and ACP domains

The six SUs are organized in pairs in three large deduced aa sequences (Fig. 1). Within each polypeptide, the end of the N-terminal SU was arbitrarily placed where the sequence C-terminal to the first ACP domain began to diverge from that of the second ACP, located toward the C-terminal end of the polypeptide. In this way, SU 1 through 6 were determined to be 1975, 1516, 1484, 1973, 1480 and 1690 aa in length, respectively. The domain order of each SU is KS, AT, KR and ACP. SU1 has additional AT and ACP domains at its N terminus, SU4 contains DH

and ER domains between AT and KR, and SU6 contains a TE at its C-terminal end (Fig. 1). We have used the different compositions of the SU as a first approximation in establishing the extent of some of the domains and confirmed the results obtained by comparison with the monofunctional PKS proteins from the *gra* (Sherman et al., 1989) and *tcm* (Bibb et al., 1989) clusters. In this way, the beginnings and ends of the KS domains could be easily assigned to the highly conserved aa motifs d(e)PiAiVgmaCR (uppercase letters refer to invariant residues) and GTNAHvIeE, respectively (Fig. 2). Comparison of the sequence of the first AT of module 1 (AT-S) to the previously aligned six other ATs clearly indicated the aa motif vfvFPGQGaQW as the likely beginning of the AT domains (Fig. 2). In a similar way, their end could be placed where AT-S is seen to diverge from the other ATs, a few residues after the highly conserved aa motif GVavdwxxa (Fig. 2). Comparison of the *eryA* ATs with the only monofunctional sequence available to us, a transacylase from *Streptomyces glaucescens* (R. Summers and C.R. Hutchinson, personal communication), confirmed the assignment of the N-terminal end and showed significant matches, albeit with two long gaps, up to the pvxLPt motif, just beyond the C-terminal end of the domains established by comparison of the *eryA* ATs.

When the segment encompassing the ACP active-site aa motif LGxDS from the first ACP of SU1 (ACP-S) was aligned with the six other *eryA* ACPs, the ends of the ACP domains appeared to coincide with the aa motif lAxhlxa,

situated 38 aa after the active-site Ser (Fig. 2). Their starts could not be easily established by this criterion, however, since conservation between ACP-S and the other six ACPs began a few aa before the active-site motif. Comparison with monofunctional ACPs from *gra* and *tcm*, however, indicated that significant matches began with the motif lAglsxxe and ended with the motif Glrlpxtlv, 15 aa before the end established by comparison of the *eryA* ACPs alone (Fig. 2). This would place the N-terminal end in all of the ACP domains, except ACP-S, 45 aa upstream from the active-site Ser.

(b) Conservation of KS, AT and ACP in multifunctional systems

The aligned KS, AT and ACP domains from the six *eryA* SUs were compared with those found in other multifunctional systems, FAS from chicken (Yuan et al., 1988; Holzer et al., 1989) and rat (Amy et al., 1989), and the PKS 6MSAS from *Penicillium patulum* (Beck et al., 1990). As expected, a higher overall degree of similarity was observed in the intra- than in the interdomain regions among the multifunctional systems analyzed (Fig. 2). The nine KS domains examined shared 71 invariant aa residues (out of 425 aa) and, considering conservative substitutions, were similar for 205/425 residues (Fig. 2). In contrast, AT and ACP domains exhibit lower conservation (Fig. 2; see also below). It is tempting to speculate that the basis for the apparently high sequence constraint in the KSs from distantly related organisms is that, in addition to catalyzing the condensation of the acyl chain with the extender unit charged on the ACP to form β-ketoacyl-ACP (Wakil, 1989), the KS is also responsible for the transacylation of the elongated acyl chain from the ACP to its own active-site Cys residue.

Invariant aa accounted for 26/345 residues in the ten AT domains examined, and approx. 30% of the ten sequences involved conservative substitutions (Fig. 2). It is noteworthy that, in addition to the segment around the signature aa sequence GHSxG, two additional segments, each containing an invariant His, are highly conserved in the ten ATs. Serine proteases are known to contain active-site Ser and His residues, distant in the primary structure, but brought into close proximity in the folded protein (Hess, 1971). Since the types of reaction carried out by ATs and serine proteases are believed to be similar (McCarthy and Hardie, 1984), the finding of invariant His in the ATs distant from the active-site Ser suggests a similarity with serine proteases also in catalytic mechanism.

The ten ACP domains examined exhibit only one invariant residue outside of the LGxDS motif. The 30-aa N-terminal segment of the ACP domains from the SU shows some conservation with mono- and multifunctional proteins, except for ACP-S (data not shown). In their inde-

pendent study of domain organization of rat FAS, Witkowski et al. (1991a) have placed the N-terminal end of the ACP domain approx. 10 aa C-terminal to the end suggested here. Thus, the ACP-S domain lacks in its N-terminal portion a segment of at least 15 aa when compared to the other ACPs. This apparent anomaly may reflect a functional difference between ACP-S and the other ACPs. According to the model proposed for 6dEB synthesis (Donadio et al., 1991), the sole role of ACP-S consists of receiving the propionyl starter unit from AT-S and of transferring it to the KS of SU1. Its function would thus be limited to acylation/deacylation, and this ACP would not be employed in carrying the β-ketoacyl chain through the appropriate processing steps, as do all other FAS and PKS ACPs known.

(c) Extent of ER, DH, KR and TE domains

Only *eryA* module 4 encodes DH and ER functions, which, to date, have only been tentatively located in FAS systems (Tsukamoto and Wakil, 1988). A 400-aa segment unique to SU4 and believed to include the ER domain (Donadio et al., 1991), was used to search the databases. Surprisingly, the best matching sequences found, aside from the rat and chicken FASs, were from structural proteins of higher eukaryotes, ζ-crystallin from guinea pig lens (Rodokanaki et al., 1989) and the membrane protein VAT-1 from *Torpedo californica* synaptic vesicles (Linial et al., 1989). The similarity of these two proteins to alcohol dehydrogenases has already been reported. In addition, the 5' end of an URF divergently transcribed from the *dnaB* gene of *Salmonella typhimurium* (Wong et al., 1988) was detected by this search. Alignment of these sequences indicated that the ER domain is likely to extend for approx. 330 aa and contains 19 invariant and 90 conserved aa residues (Fig. 3). In particular, the sequence LxHxg(a)xGGVG, proposed as the NADPH-binding site for the rat ER (Amy et al., 1989; Witkowski et al., 1991a), appears to be highly conserved in the six sequences examined. It should be noted that database searches also indicated similarity between the ER domain and alcohol dehydrogenases, but this similarity is limited mainly to the putative NADPH binding site (data not shown). Although no enzymatic role has been assigned to ζ-crystallin or VAT-1, the high similarity detected among the six sequences suggests a possible present or former role for the two monofunctional proteins in reducing double bonds that lie α,β to a carbonyl group (Piatigorsky and Wistow, 1991).

The ca. 500-aa segment defined by the end of the AT and the beginning of the ER domain in SU4 showed some similarity to the corresponding segments from rat and chicken FAS (Donadio et al., 1991), as well as with an approx. 200-aa stretch located between the AT and KR domains of 6msas (data not shown), which is believed to
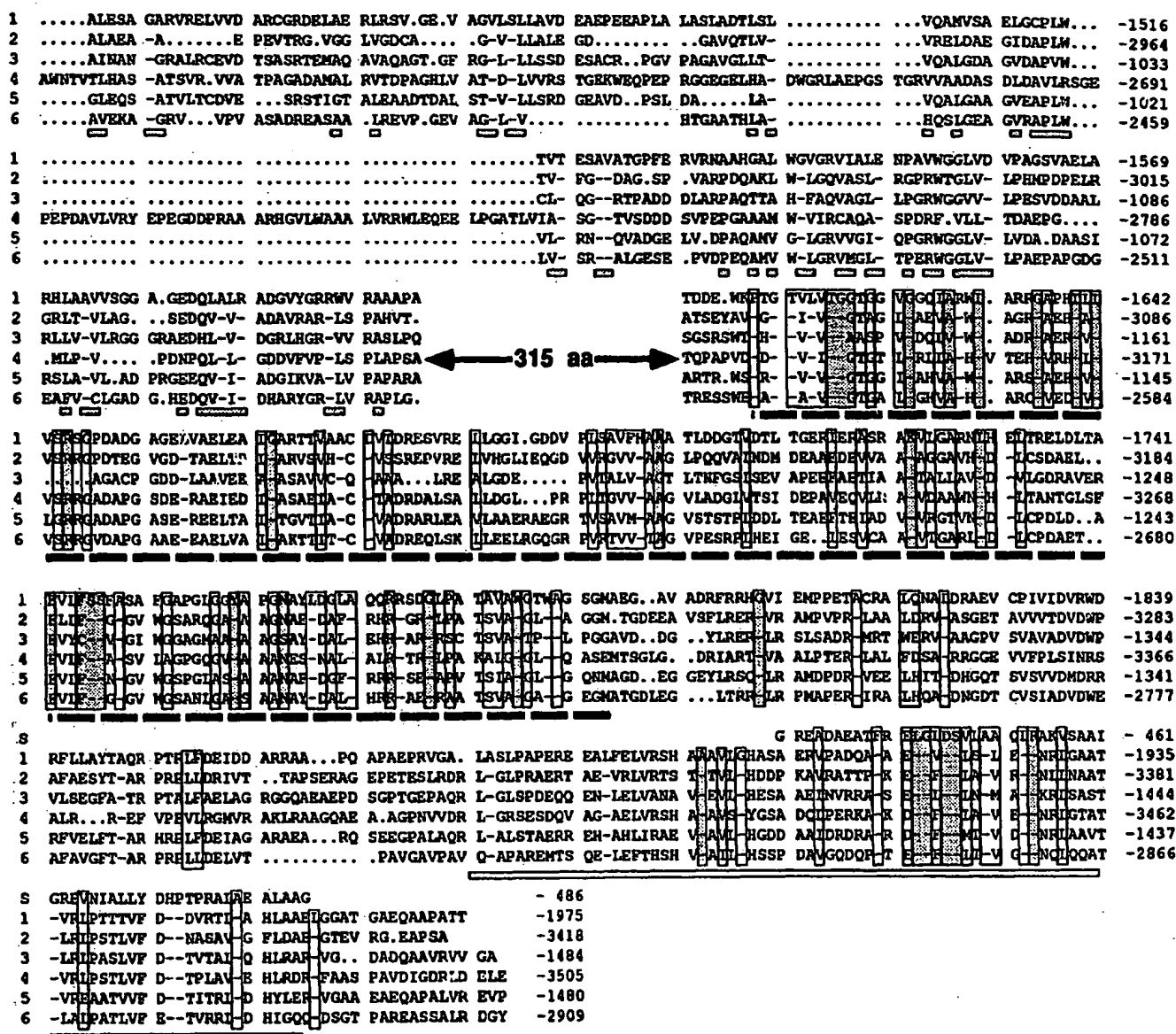
```
  1  AAPG.VAVV AMACRL.GGV STP..ME.. SEGRDAVAGL PT..GWDLDS LFHPDPTRSG TAHQRGGG.. TEATA.D.A. FG..ERERLA VD..QRL.L.     - 600
  2  APVD.-I-I GMA--L--EV DS-.RL.E.. TSGRDSAAEV -D..-VPDE LMASD....A AGTRAH-N.M AG-GD.-.A. .S..-L--M. ..QA--      -2071
  3  ELES.-I-I SMA--L-GV NT-.RL.E.. REGGETLSGF -T..-DLAR LHHPDPDNPG TSYVDK-G.. DD-AG.-.A. .S..-A--M. ..LL--       - 126
  4  ADES.-I-I GIG--F--GI G8-.QL-RV. AEGANLTTGF -A..-DIGR LYHPDPDNPG TSYVDK-G.. TD-AD.-PG. .S..-L--M. ..LM--      -1584
  5  HRAG.-I-I GMA--F-DV DS-.SE-EF. SGGGDAIAEA -A..-EPD. .......... .PDARL-G.. AA-GD.-.A. .S..-L--M. ..IM--       - 122
  6  KDAD.-I-I GMA--F-GV HN-.RL-RF. VGRGDAVTEM -T..-DLDA LFDPDPQRHG TSYSRH-A.. DG-AD.-.A. .S..-L--M. ..QV--       -1581

  1  IS.EVLR.G .P..S.CAS. .G.EVG..PQ EYGPRLAEGG EG.EG.L.IG TT.SVASG.. A.L.G.E.G. IS.T.GSS. .VAVH.AC.. .R.GESS..M    - 700
  2  TT.AL-S..  .F..T.RGSD ....V.SHQ G-ATGRPRFE DG.D..LLT. NT.S-A-.-. A-V.-.-.- .S..-.-.S .VAL-.-CGS .G.DGD..-V   -2171
  3  TS.LV-N..  .D-HS.RGTA ....V.RF G-GEDTA.AA ED.E.-SVT. VAF.-A-.-. S-T.-.-.S .S..-.-.S .VAL-.-VES .G.GESS.-V   - 225
  4  TA.AV-R..  .D..D.RGT. ....V.NGQ S-MQLLAGEA ER.D.-QGI. NS.S-L-.-. A-T.-M-.- .T..-.-.S .VGI-.-MQA .G.GEG..-L   -1684
  5  IS.AL-S..  .D..VS.RGS. ....V.GTV D-GPRPDEAP DE.L.-VGT. TA.S-A-.-. A-.I.-.-.- .T..-.-.S .AL-.-MES .D.EQ..-I   - 222
  6  TT.LF-N..  .D-HS.RGSD ....A.YQ G-GQDAVVPE DS.E.-LL.. NS.A-V-.-. A-VI.-.-.- .T..-.-.S .VAL-.-CGS .G.DGD..-V   -1680

  1  A.G..V.P.. GMLV..SRMN S.AD..RC.. .SAG.N.FG. .EGA.NL..E RIS.RRNGH PN.R.IRG.A V.S.GAS... SR.NGRA.VR V.QQA.A.SG    - 800
  2  A.VS.-AG.  EVFT.-BRQG A-S.....C-P .SDE.-.GL G.-.FVV.Q R-.-RE-R R.-GWA-S-. V.......... S-.S-VA.Q- .RR-WA.A-   -2271
  3  A.NA.-AT.  GVFVD-BRQR A-A.....S-A .GAG.-..GF S.-.VLVL.E R-.-RN-H E.-AWR-S-. I.......... S-.S-PA.R- .RG-LESC-   - 325
  4  A.VT.-SD.  YTFVD-STQR G-A.....C-A .SAR.-.AL S.-.VAALW.E P-.-AN-H Q.-AWR-S-. V.......... A-.N-PS.E- .RQ-IAAS-  -1784
  5  A.VI.-SS.  GAPT.-RSQG G-A.....C-P .SKA.-.GL A.-.AGVL.Q R-.-RE-R R.-AWR-S-. V.......... T-.S-PA.Q- .RF-LENA-   - 322
  6  A.NG.-AG.  EVFT.-BRQG G-A.....C-A .SAE.-..GF A.-.VVVL.Q R-.-RA-R Q.-GWA-S-. I.......... A-.S-VA.Q- .RK-WA.A-  -1780

  1  .G.AD.D.V. A.G.RA.L.. .IS.RA.LFEA YG..RDRE.. L..G.M.S.. G.QAAA.V.A GV.K.V..MR A..L.R.I.A SERSK.I..S SG..S.L.DEP  - 898
  2  ..DGA-VAV. .F..TA.... .V.AS.-LAT -KS-GSSG. VL..S.S.. .A.---.IV. .I-V.GL.E R.V-PN-CR G-RSG.I.-S .D.L.ADGV  -2371
  3  ..DGA-VAV. .F..... .V.AN.-LDT -RD-DADR. LW..S.S.. .A.---.VT. .I-VV.AIR N.L-AT-HV E-PTP.V.-S .V.L..AGN  - 425
  4  .F.AR-VD.. .F..... .V.AG.-IAT -QD-DR.. LR..S.S.. .A.---.VA. .I-VV.AMR H..L-RS-HA D-LSPH.I.-E .V.V..REE  -1882
  5  .RAG-VD.. .F..... .V.VH.-LST -AE-DPD. LW..S.S.. .A.---.VA. .I-AV.AIR H.EM-RT-HF D-PSPQ.I.-D I.V.VS.QA  - 422
  6  ..DGA-VAV. .F..... .V.AS.-LAT -KS-GSSG. VL..S.S.. .A.---.VA. .I-VV.GLN R.IV-PN-CR G-RSP.IF.-S .V.L.REAV  -1880

  1  EPN.RAGARP RRA.USS.F.. S.TNA.A..E EAPQVV.... .EGERVEAGD VVA.VLSAS SAEG.RACA.A RLAAHLREHP GQDPRDIAYS I.TGRAALPH    - 992
  2  RE-S.AADGV --..G-A-.V V-.A -P-EPEPVPQ PRR..MLPAT GVV-VVL-AR TGA.R--..G RIADHLAAHP GIAPADVSWT N.RA-QHFEE   -2469
  3  QP-.RRGERT --..R-A-.. V-.E -A-ERE.... .HRETTAHDG RPV-IVV-AR TTA.R--..A QIAELL.ERP DADLAGVGLG I.TT-ARHEH   - 518
  4  VP-.HAGERP --..G-S-.. V-.E -A-A...... .EQEAARTER GPL-FVL-GR 8EA.VA--..R ALAEHLRDTP ELGLTDAAWT I.TG-ARFDV  -1974
  5  RS-.HAGERP --..G-S-.. V-.E -A-EAD.... .EPEPAPDSG PV.-IVL-GR DEQ.RM--..G RIADHLAREP RNSLRDTGFT I.TR-SAWEH   - 515
  6  SP-.HAADGV --..G-A-.V V-.A -P-EPEPLPE PGPVGVLAAA NSV-VLL-AR TET.A--..R LLES..AVDD SVPLTALASA I.TG-AHLPR  -1978

  S                                              VV.V.H.CGA .N.AG.AGE.. GESRV.AAA. DACAR.FEPV I.V.TLAQV.. ........DS    - 106
  1  .A.A.PVDES A.LRVLDGI. .G.N.ADGAAV G..TSRAQQR AV.V--.-.W .E.---.AVD DTSP-.-AA. RE-AD.LEPH I.R.EVI.FL.R AEAARREQDA  -1089
  2  .A.VI.ADTA E.VHR-RAV. .D-AVVPGVVT GSASDGG... SV.V--.-.A .E.---.ARE. PV.P--.-ES. AE-DA.V.SEV A.G.SVS.EVI. .....EPRPD  -2559
  3  .A.V.ASTRE E.VRG-RE.. .R-AATADAVV EGVTEVDGRN VV.L--.-.S .A.---.GAE. SSSP--.-GK. RA-DES.MAPM Q.D.VKVS.DV.. .....RQAPG   - 612
  4  .A.VI.GDDRA G.VC.AE-DA. .E-RPSADAVA PVTSAP..RK PV.V--.-.A .A.---.ARD. ESSE--.-ES. SR-AE.I.SPH T.D.VKLL.DVV. .....RGDGG  -2066
  5  .VV.G.DRD D.ALAG-RAV. .D-RIADRTAT G..QARTRRG VA.V--.-.A .E.---.ARD. RESQ--.-DS. RD-ER.A.IAPH V.D.WSLT.DLL. ........SG  - 603
  6  .A.L.AGDHE Q.RGQ-RAV. .E-VAAPGATT GTASAGG... VV.V--.-.A .F.---.ARG. SV.P--.-ES. AE-DA.V.SEV A.G.SASE.VI. .....EQRPD  -2068

  S  PEQSRR.EV. QP.LFAV.I. .AA.W.S.GV .R.AVVGHS. GE.A.RH.CG AAG.ADA.A. A.I.WSRE.IP I.VGN.D.A.V ALSADEIEPR IA..RWDDD.    - 204
  1  ALSTE.-D-V --VM-AVMV- -.-SM-RA.-N E-A--.--.Q D.-.-V-A. ALSLDDA.RV VA.IRSRV.IAT .PGN.G-.A.I A...APAGEV RA..RIGDR.  -1184
  2  APSLE.-D-V --VL-AVMV- -RL-RA.A. V-S.-.--.Q I.-.-V-A. ALSLEDGN.RV VA.RRSRA.VRA VA.GR.S-.SV. RGGRSDVEKL LADDSWTGR.  -2659
  3  APGLD.-D-V --VL-AVMV- -EL-RS.. E-A--V--.Q I.-.-V-A. ALT.LEDA.R.I VV.GRSRI.MRS I.SGE.G-.AV. ALGEAAVRER LR..PWQDR.  - 710
  4  PDPHE.-D-I --VL-S.MV- -EL-RA.. T-A--V--.Q I.-.-H-A. ALS.LERA.R.V VA.IRSQV.LRE L.DD.Q.G-.SV. GASRDELETV LA..RWDGR.  -2164
  5  ARPLD.-D-V --AL-RVMV- -AL-RS.N E-A--.--.Q I.-.-V-A. ALT.LEDA.N.I VA.VRSRV.IAR L.GG.Q.G-.SF. GLGTEQAAER IG..RFAGA.  - 701
  6  APSLE.-D-V --VL-S.MV- -RL-GA.N E-S--.--.Q I.-.-H-A. VLS.LEDG.VR.V VA.IR.KA.LRA I.AG.R.G-.SU. AAPGERARAL IA..PWEDR.  -2166

  S  VI.G.V.GP.. VIL.TGSPEE.V ARR.QELSAE G.RAQV.NVS .A.A.GAC.DD I.AEG.RSA.A W.E.APGGSEVP FY.A.I.TGGAV .DTRELVAD. WR..SERL..R    - 303
  1  EI.A-G-RS. VVV.A-DSDE.. DRLVASCTTE Q.RAKRI.A-D M.S.SSH.ET I.RDA.HAE.G EDF.HPL.GFV PF.-.IVTGRW TQPDELDAG. -YRN.I-RT-.R  -1284
  2  EV.A-G-D. VVV.F-DAQR.A REFLEYCEGV G.RARA.IP-D M.S.TAH.EP VRDE.LVQA.IA .G.ITPRRAEV PF.-.ILTGDF LDGTELDAG. -YRN.I-HF-.E  -2758
  3  SV.A-G-RS. VVV.S-EPGR.. RAFSEDCAAE G.RVRD.ID-D M.S.PQ.IER VRBE.IL.EIT G.DIAPRPARV TF.-.IVESRS MDGTELDAR. -YRN.I-ET-.R  - 809
  4  AV.A-G-G.. SVV.A-PTAE.. DEFFAEAEAR E.KPRR.IA-R M.S.SPE.VAR .EDR.IAAE.G ..TITAVRGSV P.H-.IV.GEV IDTSANDAS. -YRN.I-RF-.L  -2263
  5  SI.S-G-RS. VVV.A-ESGR.. DELIAECEAB Q.ITARR.IP-D M.S.SPQ.VES IREE.LLTE.IA .GISPVSADV A.M-.ITGQP IDTATMDTA. -YAN.I-EQ-.R  - 800
  6  SV.A-L-SS. VVV.S-DPEA.. AELVARCEDE Q.VRAKT.IP-D L.S.SRH.VEE IRET.ILAD.D ..QISARRAAI P.V-.ILH.GER RDGADMGPR. -YDN.I-SQ-.R  -2265

  S  F.D.A.RSAL.E V.GPG.F.A.. .R.PVL.AA.Q QTLDAE...G SSAAV.P.I.Q RGQGG.RR.L LAAAQAFTG. VA.VDWTAAYD DVGPNPAL                  - 388
  1  AD.-VRA.AE C-YRT-I-V- A.-.I.TA.IE EIGDG...SG ADLS.H.I.S-R .GDGSLAD.G EALSR-FA. .A.VDWESVHL GTGARRVPLP TY.P.ORER..    -1381
  2  HS.-VQALTD C-YAT-I-V- P.-.VI.ASSVQ ETLDD...AE SDAA.V.GT.E .DAGDADR.L TALAD-HT. .A.VDWEAVL. .GRAGLVD-- G.--QGK-.    -2853
  3  AD.-VTRLAE S-YDR-I-V- E.-.VV.QA.E EAVEEADGAE .DAV.VG.S-H .DGGDLSA.E RSMAT-HVS. .I.RWDVAL. .PGAAPFA-- T.--ORK-Y.    - 906
  4  EQ.-VRGLVE C-FDT-V-V- B.-.VI.LMA.E ETAEH...AG AEVTQ.V.PT.-E .EQSGPHEF.L RNLLR-HV. .A.DLRPAV. .AGGRPAE-- T.--EHQ-.    -2358
  5  QD.-TRQLAE A-FDA-V-V- B.-.VI.TVG.IE ATLDSALPAD AGAC.V.GT.-R .DRGGIAD.H TALGE-YA. .A.VDWSPAF. .ADARPVE-- V.--QRQ-Y.    - 898
  6  DE.-VSAA.VA D-HAT-V.M- F.-.VI.TAA.VQ EI........A ADAVA.IGS.-H .DTAE.EH.I AELAR-HV. .A.VDWRNVE. .PAAPPVA-- N.--EPQ-Y.    -2355

  1  LEPKPVARRS TE               VDEVSALRY. ........RI EWRPTGA... ........GE. ARLDGTWLVA KYAGTADETS TAA.RE....  -1439
  2  L....LPDRT TP               R-ELDGWF-. ........RV D-TEVPR... ........SE. AAIRGRW-VV VPEGHEEDGW TVEVRS....  -2908
  3  LQPAAPAAA. ..               S-ELA...-. ........RV S-TPIEK... ........PES GNLDGDW-VV T.PLISPE.W TEMLCE....  - 957
  4  PRPHRPADVS AL  ◄— 152 aa —► Y-GLAEQG-E YGPSFQALRA A-RKDDSVYA EVSIAADEEG YAFHPVL-DA VAQTLSLGAL GEPGGGKLPF  -2592
  5  L....PIPTGG RA              R-EDDDWR-. ........QV V-REAEH... ........ES ASLAGRV-LV TGPGVPSE.L SDAIRS....  - 952
  6  L....APE.. .V               S-QLADSR-. ........RV D-RPL.A... ........TT. VDLEGGF-V. ..HGSAPESL TS........  -2399
```

```
1   .....ALESA GARVRELVVD ARCGRDELAE RLRSV.GE.V AGVLSLLAVD EAEPEEAPLA LASLADTLSL .......... ...VQAMVSA ELGCPLW...  -1516
2   .....ALAEA -A.......E PEVTRG.VGG LVGDCA.... .G-V-LLALE GD........ ..GAVQTLV- .......... ...VRELDAE GIDAPLW...  -2964
3   .....AIHAN -GRALRCEVD TSASRTEMAQ AVAQAGT.GF RG-L-LLSSD ESACR..PGV PAGAVGLLT- .......... ...VQALGDA GVDAPVW...  -1033
4   AWNTVTLHAS -ATSVR.VVA TPAGADAMAL RVTDPAGHLV AT-D-LVVRS TGEKWEQPEP RGGEGELHA- DWGRLAEPGS TGRVVAADAS DLDAVLRSGE  -2691
5   .....GLEQS -ATVLTCDVE ...SRSTIGT ALEAADTDAL ST-V-LLSRD GEAVD..PSL DA.....LA- .......... ...VQALGAA GVEAPLW...  -1021
6   .....AVEKA -GRV...VPV ASADREASAA .LREVP.GEV AG-L-V.... .......... HTGAATHLA- .......... ...HQSLGEA GVRAPLW...  -2459

1   .......... .......... .......... .......... .......... .......TVT ESAVATGPFE RVRNAAHGAL WGVGRVIALE NPAVWGGLVD VPAGSVAELA  -1569
2   .......... .......... .......... .......... .......... .......TV- FG--DAG.SP .VARPDQAKL W-LGQVASL- RGPRWTGLV- LPHNPDPELR  -3015
3   .......... .......... .......... .......... .......... .......CL- QG--RTPADD DLARPAQTTA H-FAQVAGL- LPGRWGGVV- LPESVDDAAL  -1086
4   PEPDAVLVRY EPEGDDPRAA ARHGVLWAAA LVRRWLEQEE LPGATLVIA- SG--TVSDDD SVPEPGAAAM W-VIRCAQA- SPDRF.VLL- TDAEPG....  -2786
5   .......... .......... .......... .......... .......... .......VL- RN--QVADGE LV.DPAQAMV G-LGRVVGI- QPGRWGGLV- LVDA.DAASI  -1072
6   .......... .......... .......... .......... .......... .......LV- SR--ALGESE .PVDPEQAMV W-LGRVMGL- TPERWGGLV- LPAEPAPGDG  -2511

1   RHLAAVVSGG A.GEDQLALR ADGVYGRRWV RAAAPA           TDDE.W??TG  [TVLV??G??G ?GG??ARN?. ARF?RPH???]  -1642
2   GRLT-VLAG. ..SEDQV-V- ADAVRAR-LS PAHVT.           ATSEYAV-G-  [-I-V??TAG ??AE?A-W?. AGH?AE???]    -3086
3   RLLV-VLRGG GRAEDHL-V- DGRLHGR-VV RASLPQ           SGSRSWT-H-  [-V-V???SP ??DI?-W?. ADR?AE??V]     -1161
4   .MLP-V.... .PDNPQL-L- GDDVFVP-LS PLAPSA  ◄——315 aa——►  TQPAPV?-D-  [-V-I???GG ?RI?A-H?V TER?R?H?]  -3171
5   RSLA-VL.AD PRGEEQV-I- ADGIKVA-LV PAPARA           ARTR.NS-R-  [-V-V??G?G ??AH?A-W?. ARS?AE??]     -1145
6   EAFV-CLGAD G.HEDQV-I- DHARYGR-LV RAPLG.           TRESSW?-A-  [-A-V??G?A ??GH?A-H?. ARC?VED??]    -2584

1   V?R?QPDADG AGE?VAELEA ??ARTI?AAC ?V?DRESVRE ?LGGI.GDDV ??SAVFH??A TLDDGT?DTL TGER?ER?SR A??IGARN?? ??TRELDLTA  -1741
2   V?R?PDTEG  VGD-TAELI? ?-ARVSVH-C ?VSSREPVRE ?VHGLIEQGD V??GVV-??G LPQQVA?NDM DEA?EDE?VA A??GGAVH?D -?CSDAEL..  -3184
3   ??.?AGACP  GDD-LAAVEE ?-ASA?V?C-Q ???A...LRE ?LGDE..... ?V?ALV-?QT LTN?GS?SEV APEE?AE?LA A??ALIAV?D -?LGDRAVER  -1248
4   V?R?ADAPG  SDE-RAEIED ?-ASAE?A-C ?W?DRDALSA ?LLDGL...PR ?I?GVV-?AG VLADGI?TSI DEPA?EQVL? A??DAA?N-H -?TANTGLSF  -3268
5   L?R?ADAPG  ASE-REELTA ?-TGVT?A-C ?W?DRARLEA ?VLAAERAEGR ?VSAVM-?AG VSTSTF?DDL TEAE?TE?AD V??RGTVN?D -?LCPDLD..A -1243
6   V?R?GVDAPG AAE-EAELVA ?-?ART??T-C ?W?DREQLSK ?LLEELRGQGR ?V?TVV-??G VPESRF?HEI GE..?ES?CA A??VTGARL?D -?LCPDAET.. -2680

1   ?V??????SA ??GAPGI???A ??NAYLQ?LA ?Q??RSDQ?PA ?AVA?G?WAG SGHAEG..AV ADRFRRH?VI ENPPETA?CRA L?NA?DRAEV CPIVIDVRWD  -1839
2   ?LI?-?-GV  W?SARQ?A-A ?A?NAE-?AF- RH?-GR-L?A ?SVA-?GL-?A GGN.TGDEEA VSFLREF-VR AMPVPR-LAA L?RV-ASGET AVVVTDVDWP  -3283
3   ?VI?-?-GI  W?GAGHAA-A ?A?SAV-D?L- EH?-AF-?SC ?SVA-?IP-L PGGAVD..DG ..YLREF-LR SLSADR-MRT WERV-AAGPV SVAVADVDWP  -1344
4   ?VI?-?-SV  IA?GPGQ?V-A ?A?NES-NAL- AIA-TR-L?A RALG-?GL-Q ASEMTSGLG. .DRIART-VA ALPTER-LAL F?DSA-RRGGE VVFPLSINRS -3366
5   ?VI?-?-GV  W?SPGIA?-A ?R?N-SE-?A?V ?SIA-?GL-L Q?MAGD..EG GEYLRSG-LR AMDPDR-VEE L?IT-DHGQT SVSVVDMDRR -1341
6   ?VI?-?-GV  W?SAN?GA?S ?A?NAY-D?L- HF?-AF-R?A ?SVA-?GA-?G EGMATGDLEG ...LTRR?-LR PMAPER-IRA L?QA-DNGDT CVSIADVDWE  -2777

                                                                      G  RE?ADAEAT?R  [?E?G??DSV?AA  Q??AK?USAAI]  -  461
1   RFLLAYTAQR PT?LF?EIDD ARRAA...PQ APAEPRVGA. LASLPAPERE EALFELVRSH A?A??GHASA E??PADQA-A E?-?V-??L?-L E?-NR??GAAT  -1935
2   AFAESYT-AR PR?L?DRIVT ..TAPSERAG EPETESLRDR L-GLPRAERT AE-VRLVRTS T?TV?-HDDP KA?RATTH-K B?-??-?A?-V R?-NI??NAAT  -3381
3   VLSEGFA-TR PT?LF?AELAG RGGQAEAEPD SGPTGEPAQR L-GLSPDEQQ EN-LELVAHA V??V?-BESA AE?NVRRA-S E?-??-??A-M A?-KR??SAST  -1444
4   ALR...R-EF VP?VL?RGHVR AKLRAAGQAE A.AGPNVVDR L-GRSESDQV AG-AELVRSH A??A?S-YGSA DQ??PERKA-K D?-??-??A-V E?-NR??GTAT  -3462
5   RFVELFT-AR HR?LF?EIAG ARAEA...RQ SEEGPALAQR L-ALSTAERR EH-AHLIRAE V??A?L-HGDD AA?DRDRA-R E?-??-??L-V D?-NR??AAVT  -1437
6   AFAVGFT-AR PR?L?DELVT .......... .PAVGAVPAV Q-APAREHTS QE-LEFTHSH V??A?L-HSSP DA?GQDQF?-T E?-??-??L-V G?-NQ??QQAT  -2866

S   GRE?NIALLY DHPTPRAI?E ALAAG                 - 486
1   -VR?PTTTVF D--DVRTI?A HLAAE?GGAT GAEQAAPATT  -1975
2   -LR?PSTLVF D--NASAV?G FLDAE?GTEV RG.EAPSA    -3418
3   -LR?PASLVF D--TVTAI?Q HLRAF?VG.. DADQAAVRVV GA  -1484
4   -VR?PSTLVF D--TPLAV?E HLRDF?FAAS PAVDIGDR?D ELE  -3505
5   -VR?AATVVF D--TITRI?D HYLEF?VGAA EAEQAPALVR EVP  -1480
6   -LR?PATLVF E--TVRRI?D HIGQ??DSGT PAREASSALR DGY  -2909
```

Fig. 2. Alignments of the six *eryA* SU. The symbols on the left margin refer to the particular SU, with S referring to AT-S or ACP-S, as appropriate. Numbers on the right margin refer to the aa sequence position at the end of each row in EryAI (for 1, 2 and S on left), in EryAII (for 3 and 4) and in EryAIII (for 5 and 6). Sequences for EryAI and for EryAII and -III are from GenBank, accession Nos. M63676 and M63677, respectively. Invariant aa residues in the six SU are marked by dashes. Dots refer to computer-introduced gaps to maximize alignments. Shaded boxes refer to aa residues invariant in the six (or seven) sequences from the SU, as well as chicken FAS (Holzer et al., 1989; Yuan et al., 1988), rat FAS (Amy et al., 1989) and 6MSAS (Beck et al., 1990). Open boxes refer to conservative substitutions or invariant residues in all but one sequence. The N terminus of chicken FAS is assumed to precede the published sequence (Holzer et al., 1989), as recently reported (Witkowski et al., 1991a). The KR of SU3, when it deviates from the other eight sequences, is ignored for boxing purposes. The extent of each domain is indicated by underlining of the sequences with solid black bars, short, heavy dashes, long, heavy dashes, and open bars, representing the KS, AT, KR and ACP domains, respectively. The two arrows mark the extra segments of 152 and 315 aa present in SU4, which are presented in Figs. 4 and 3, respectively. The shaded bars under the sequences in the region comprised between the two arrows indicate invariant and conservative substitutions among the six SU. Computer-assisted sequence analyses were performed using the University of Wisconsin GCG programs (Devereux et al., 1984). Sequences were examined pairwise using COMPARE/DOTPLOT. Multiple sequence alignments were performed using PILEUP, with a gap weight of 3.0 and a gap length weight of 0.1. The sequences of the six SU were initially aligned. Subsequently, the segments corresponding to the first AT and ACP of SU1 (AT-S and ACP-S, respectively) were individually aligned with the other six AT and ACP domains, respectively. Finally, the region of SU4 between the DH and ER domains (section c) and those from SU 1, 2, 3, 5 and 6 between the AT and KR domains were separately aligned. The three alignments generated in this way were manually combined using LINEUP. For comparing the six *eryA* SU with the other multifunctional systems examined, chicken FAS, rat FAS and 6MSAS, PILEUP was run using one sequence from each group (one SU, one FAS and 6msas) or with all nine sequences, with similar results. The DH-ER interdomain region of SU4 and the AT-KR interdomain regions from the other five SU, when compared with the other multifunctional systems, gave substantially different alignments upon changing of the PILEUP parameters. These segments have thus been ignored for boxing purposes.
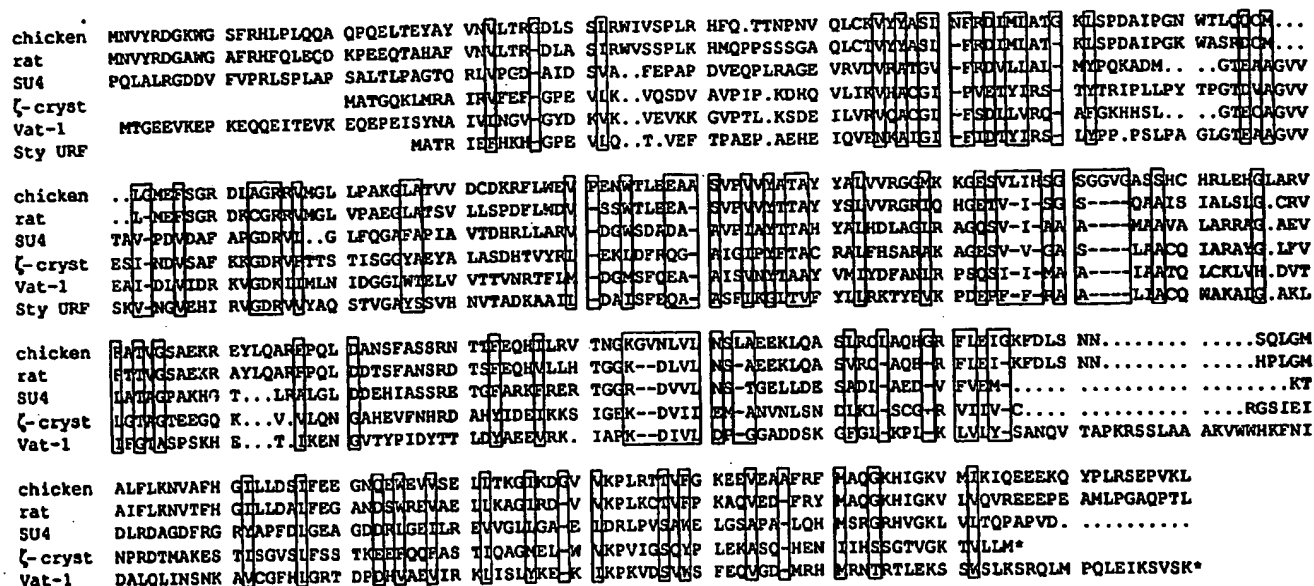
```
chicken  MNVYRDGKWG SFRHLPLQQA QPQELTEYAY VNVLTRGDLS SIRWIVSPLR HFQ.TTNPNV QLCKVYYASI NFRDIMIATG KISPDAIPGN WTLQCM...
rat      MNVYRDGAWG AFRHFQLECD KPEEQTAHAF VNVLTR-DLA SIRWVSSPLK HMQPPSSSGA QLCTVYYASI -FRDIMIAT- KISPDAIPGK WASHDCM...
SU4      PQLALRGDDV FVPRLSPLAP SALTLPAGTQ RIVPGD-AID SVA..FEPAP DVEQPLRAGE VRVDVRATGV -FRDVLIAI- MVPQKADM.. ..GTEHAGVV
ζ-cryst             MATGQKLMRA IRMFEF-GPE VLK..VQSDV AVPIP.KDKQ VLIRVHACGI -PVETYIRS- TVTRIPLLPY TPGTVVAGVV
Vat-1    MTGEEVKEP KEQQEITEVK EQEPEISYNA IVDNGV-GYD KVK..VEVKK GVPTL.KSDE ILVRVQACGI -FSDILVRQ- AFGKHHSL.. ..GTEDAGVV
Sty URF                      MATR IEEHKH-GPE VLQ..T.VEF TPAEP.AEHE IQVENRAIGI -FIDTYIRS- LIPP.PSLPA GLGTEHAGVV
```

```
chicken  ..LQMEFSGR DIAGRRVMGL LPAKGLATVV DCDKRFLWEV PENRTLBEAA SVFHVMITAY YRLVVRGGMK KGBSVLIHSG SGGVGASSHC HRLEHGLARV
rat      ..L-MEFSGR DRCGRRVMGL VPAEGLATSV LLSPDFLMDV -SSWTLBEA- SVFHVMITAY YSLVVRGRDQ HGBHV-I-SG S----QAAIS IALSIG.CRV
SU4      TAV-PDVDAF APGDRVI..G LFQGAFAPIA VTDHRLLARM -DGWSDADA- AVFLAMITAH YRLHDLAGIR AGGSV-I-AA A----MAAVA LARRAG.AEV
ζ-cryst  ESI-NDMSAF RKGDRVTTS TISGGYAEYA LASDHTVYRD -EKIDFRQG- AIGIPMFTAC RALFHSABHK AGBSV-V-GA S----LAACQ IARAVG.LFV
Vat-1    EAI-DIMIDR KVGDRIIMLN IDGGIWTELV VTTVNRTFIM -DGMSFQEA- AISVRMTAAY VMIYDFANIR PSQSI-I-MA A----IAATQ LCKLVH.DVT
Sty URF  SKV-NGWEHI RVGDRVMYAQ STVGAYSSVH NVTADKAAIU -DAISFRQA- RSFLRGLTVF YILRKTYEWR PDEFF-F-HA A----LAACQ MAKAIG.AKL
```

```
chicken  RAIWGSAEKR EYLQARFPQL DANSFAS8RN TIFEQHILRV TNGKGVNLVI NSLAEEKLQA SIRQIAQHGR FLEIGKFDLS NN........ .....SQLGM
rat      RTIVGSAEKR AYLQARFPQL DDTSFANSRD TSFEQHVLLH TGGK--DLVI NS-AEEKLQA SVRQ-AQE-R FLEI-KFDLS NN........ .....HPLGM
SU4      LATIGPAKHG T...LRALGL DDEHIASSRB TGFARKFRER TGGR--DVVI NS-TGELLDE SADI-AED-V FVEM-..... .......... ........KT
ζ-cryst  LGTVGTEEGQ K...V.NLQN GAHEVFNHRD AHYIDEIKKS IGEK--DVII DM-ANVNLSN DIKI-SCG-R VIIV-C.... .......... ....RGSIEI
Vat-1    IFGTASPSKH E...T.IKEN GVTYPIDYTT LDYAEEVRK. IAFK--DIVI DF-GGADDSK GFGI-KPI-K LVILY-SANQV TAPKRSSLAA AKVWWHKFNI
```

```
chicken  ALFLKNVAFH GILLDSIFEE GNQEWEVWSE IDTKGIKDGV NKPLRTIVFG KEEVEAAFRF HAQGKHIGKV MIKIQEEEKQ YPLRSEPVKL
rat      AIFLKNVTFH GILLDAIFEG ANDSWRFVAE IIKAGIRD-V NKPLKCTVFP KAQVED-FRY MAQGKHIGKV INQVREEEPE AMLPGAQPTL
SU4      DLRDAGDFRG RVAPFDIGEA GDDRIGEILR EVVGILGA-E IDRLPVSAVE LGSAPA-LQH MSRGRHVGKL VLTQPAPVD. ..........
ζ-cryst  NPRDTMAKES TISGVSLFSS TKEEHQQFAS TIQAGNEL-W NKPVIGSQVP LEKASQ-HEN IIHSSGTVGK TVLLM*
Vat-1    DALQLINSNK AWCGFHLGRT DPDHVAEVIR KIISLMKE-K LKPKVDSWHS FEQVGD-MRH MRNTRTLEKS SMSLKSRQLM PQLEIKSVSK*
```

Fig. 3. Putative enoyl reductase domains. The segment from EryAII comprised between aa 2832 and 3138 was employed to screen the GenBank and EMBL (48 285 sequences) and the Swissprot (20 722 sequences) databases using the programs TFASTA and FASTA, respectively. Sequences showing significant matches were aligned using PILEUP. Invariant aa residues are represented by dashes; dots refer to inserted gaps. Conserved regions among all the sequences are boxed. An asterisk indicates the stop codon in the corresponding gene. Sequences as indicated: chicken, FAS from chicken, aa 1483–1853; rat, FAS from rat, aa 1496–1866; SU4, *S. erythraea* EryAII, aa 2795–3138; ζ-cryst, guinea pig lens crystalline, complete sequence (Rodokanaki et al., 1989); VAT-1, membrane protein from *Torpedo californica* cholinergic synaptic vesicles, complete sequence (Linial et al., 1989); Sty URF, *Salmonella typhimurium* unidentified open reading frame divergent from *dnaB*, translated from GenBank J03390 (Wong et al., 1988). Note that only the 5' end of this sequence is available, with the resulting polypeptide ending as KALGAKL[168].

contain a DH function (Beck et al., 1990). The alignment of these four sequences (Fig. 4) indicates that significant homology is limited to an approx. 150-aa segment. Within it, the invariant HxxxGxxxxP motif is embedded in a 25-aa segment with a high degree of conservative substitutions, involving mostly hydrophobic residues. The finding of an invariant His residue in the most conserved region among the four sequences is consistent with the proposed role for a His as the active-site residue in the *E. coli* β-hydroxydecanoyl thioester dehydrase (Bloch, 1971). The corresponding gene encodes the active-site His in the sequence HFIGDPVMP[78] (Cronan et al., 1988), where insertion of a gap between I and G would conform this sequence to the proposed consensus. The HxxxGxxxxP motif

is also found at a single position in the *S. cerevisiae* FAS1 sequence as HLSHGVKMIP[1057], approximately where the DH domain has been tentatively placed (Schweizer et al., 1986; Chirala et al., 1987). These observations suggest that the DH domain in multifunctional FAS and PKS systems is relatively short, extending for approx. 140–170 aa, consistent with the 170-aa size of the *E. coli* enzyme, and point to a specific His as one of the active-site residues involved in catalysis. The same His has been independently proposed by P.F. Leadlay (personal communication) as the catalytic residue in the DH domain of EryAII. It should also be noted that the two animal FASs and SU4 also share the motif GYxYGPxFQ, approx. 110 aa after the proposed active-site His, whereas 6msas does not. The

```
chicken  DHSQWDVPK AED...FPSG SKGSASVSVY NIDVSPDSPD HVLVGHQIDG RVLYPATGYI VIARHLARS LGMVMEDTAV MFEEVIIHQA TIIPKKGSTQ
rat      DHSQWDIPV AED...FPNG S.SSSSAT-Y NIDASSESPD HVLVD-QID- RVLF-GIGYI YIVWKFLARS LSLSUEETH- VFENVIEHQA TIIPRTGTVP
SU4      ADVSRIGVRG AEHPLLLAAV DVFGHGGA-F TGRLSTDE.Q FWLAE-NVG- RVLV-GSVLV DIALAA.GED VGLPVIEEL- IQRPLVLAGA GRILRMSVGA
6msas    PLHTQITHDV EKHTLLGQRI PVFGTDTY-Y TTRLDNT.K PFPGS-RVH- TEIV-AKGLI NTFIRG.TGG ...QMLQNV- LRVPVAUINAB RSV...QVVV
```

```
chicken  IEVRIMPASH SFEVSGNGND AVSGKISIDE NDALKNFHNQ IRDFQSQARV TAKSGLLMED VYQEIHL... ......RGY NVGPTFGGV. ..........
rat      IEVRLLEASH AFEVSDSGNL IVSGKVYQME DPDSKLFDH. JFEVPIPAFS ESVSRLTQGE VYKEIRL... ......R-Y DVGPHFGGV. ..........
SU4      FDESGRRTID VHAAEDVADI ADAQ....MS QHATGTLAQG VAAGPRDTDQ WPPEDAVRIP IDDHKD.... ...GLAEQ-Y EVGPSFCALR AAWRKDDSVY
6msas    QQDQVKVVSR LIPSEPSQLD DDHS....MV THTTAYYDRK VAGSEDRIDF AAVKSRLVTK IIADNESIDYL DKVGVSAM-E PH....AVT EHYRNDKEML
```

Fig. 4. Putative dehydratase domains. The approx. 500-aa segments from EryAII, chicken FAS, rat FAS and 6msas suspected to contain the DH domain (see section e) were aligned using PILEUP. Only the portion showing significant matches is represented. For abbreviations and symbols, refer to Fig. 3. The three blackened squares denote the putative active-site motif HxxxGxxxxP. The shaded bar denotes the highly conserved region common only to the two FAS sequences and SU4. Sequences: chicken, FAS from chicken, aa 812–987; rat, FAS from rat, aa 837–1009; SU4, *S. erythraea* EryAII, aa 2365–2551; 6msas, *Penicillium patulum* 6MSAS, aa 914–1096.

reason for this difference and a possible role for this motif are at present unknown. No significant matches were de- tected by database searching with the proposed DH do- main or by comparison with known aa dehydratases.

All six *eryA* SU contain a segment corresponding to a KR domain, although the KR domain of SU3 is believed to be non-functional (Donadio et al., 1991). The beginning of the KR domains is likely to coincide with the region following the ER domain where SU4 realigns with the other five SU. This location was matched in the N-terminal por- tion of the monofunctional KRs involved in actinorhodin (Hallam et al., 1988) and granaticin (Sherman et al., 1989) synthesis. Thus, the *ery* KR domains are likely to start with the PxGTvLv motif, just upstream from the putative NADPH-binding site (Fig. 2). Since in all multifunctional FAS and PKS systems the KR is always followed by an ACP, the end of the KR domains was placed approx. 190 aa after the NADPH-binding site, where conservation among the nine sequences examined began to decline (Fig. 2). This interpretation results in the separation of the KR and ACP domains of 90–100 aa in the PKS systems and of 60 aa in the FAS systems.

The C-terminal end of SU6 contains a TE domain. This domain was compared with the corresponding domains from the two FAS sequences, with monofunctional thioesterases from rat (Randhawa and Smith, 1987; Saf- ford et al., 1987) and duck (Poulose et al., 1985), and with the TE-like ORF downstream from *eryF* (Weber et al., 1991). The alignments (Fig. 5) indicate that the TE domain in *eryA* extends for approx. 230 aa, and includes, in addi-

tion to the invariant GxSxG motif common to ATs and serine proteases, the GdH motif found near the C-terminal end, which has been shown by site-directed mutagenesis to be essential for activity (Witkowski et al., 1991b). Overall, little similarity was detected among the six TEs analyzed, which may be related to the three different classes of sub- strates recognized by these enzymes (Wakil, 1989; Cortes et al., 1990; Donadio et al., 1991). This is exemplified by the low similarity between the TE domain of rat FAS and the short chain TE from the same organism, and between the TE of SU6 and the other TE-like sequence from the *ery* cluster, although the role of the latter has not yet been determined.

(d) Inter- and extradomain regions

The overall domain organization of the three *eryA*-encoded polypeptides is summarized in Fig. 6. It can be seen that the largest interdomain regions are the five seg- ments between AT and KR, and the one between DH and ER in EryAII. When the AT-KR interdomain regions from SU 1, 2, 3, 5 and 6 were compared with the region from SU4 containing the DH and ER domains, some similarity could be detected under relatively stringent conditions (data not shown). Computer-generated alignment of these six segments indicated that these regions can be best accom- modated after accounting for two insertions in SU4, the first of 152 aa, and the second of 315 aa (Fig. 2). These two insertions correspond very closely to the DH and ER do- mains, respectively, as determined above. In the 200-aa segment which joins the DH and ER domains in SU4, a

Fig. 5. Thioesterase domains. The six TEs were compared using PILEUP. See legend to Fig. 3 for symbols. Sequences: rat, FAS from rat, aa 2209– 2505; chicken, FAS from chicken, aa 2193–2497; rat SC, short-chain TE from rat, complete sequence (Randhawa and Smith, 1987; Safford et al., 1987); duck SC, short-chain TE from duck, complete sequence (Poulose et al., 1985); eryORF, downstream from *S. erythraea eryF* (Weber et al., 1991; GenBank accession No. M54983); SU6, *S. erythraea* EryAIII, aa 2927–3170.
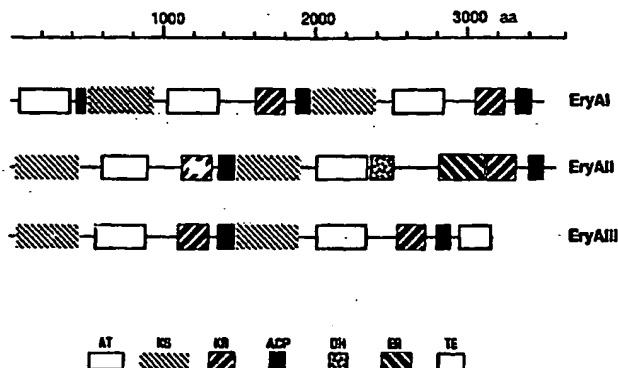
Fig. 6. Domain organization in 6dEB synthase. Each domain is represented by a rectangle of different filling as shown, whose length is proportional to the length of the domain. Note the partial filling of the first KR of EryAII, which denotes an inactive KR (Donadio et al., 1991).

stretch of approx. 80 aa appears to be fairly well conserved among the six SU (Fig. 2). The enzyme from *S. erythraea* AKR5, which has been deleted of this 80-aa segment along with the KR domain C-terminal to it in SU5, retains active SU5 and SU6, as judged by the ability to produce a significant amount of the 6dEB analog lacking the hydroxyl group introduced during synthesis step 5 (Donadio et al., 1991; Fig. 1). Whilst Witkowski et al. (1991a) have speculated that the long DH-ER interdomain segment is involved in facilitating protein–protein interactions in the dimeric FAS enzyme, our results indicate that the presence of at least a portion of this segment is not absolutely required for 6dEB PKS function, although the level of activity of the altered enzyme could not be measured directly. That the deleted segment is important for KR activity is improbable, since very little homology was detected with other multifunctional systems in this region (data not shown).

An additional feature of the *eryA*-encoded polypeptides is the presence of extra N-terminal and C-terminal tails extending significantly beyond the domain limits (Fig. 6). The N termini of polypeptides EryAII and EryAIII contain segments of 26 and 33 aa preceding the KS domains of SU3 and SU5, respectively, and the C termini of EryAI and EryAII contain stretches of 69 and 63 aa following the ACP domains of SU2 and SU4, respectively. The other multifunctional systems examined do not contain extra tails of such lengths. In 6msas a 28-aa segment precedes the KS domain, but the ACP domain is followed by a segment of only 6 aa at the C terminus of the polypeptide. The KS domain of rat FAS starts at the N terminus of the protein. It is tempting to speculate that these additional segments in the *eryA*-encoded polypeptides may play a role in facilitating the correct intermolecular transfer of the growing acyl chain, such as from SU2 in EryAI to SU3 in EryAIII, either by enabling specific protein–protein interactions, or

by properly positioning the polypeptides on some cellular structure.

(e) Evolution of the modules

S. erythraea contains a Type-I PKS and, most likely, a Type-II FAS system (Revill and Leadlay, 1991). The evolutionary origin of these two systems can be understood by comparison of similar enzymatic functions belonging to a Type-I or Type-II system from different sources, as exemplified for the 21 ACPs presented in the dendrogram in Fig. 7. The *eryA* ACPs are closely related to each other, except for ACP-S, which, as described above, does not function as all other known ACPs and is less related to the other SU ACPs than the ACP from 6msas. Nonetheless, the Type-I PKS ACPs appear to be clustered together indicating greater overall similarity amongst each other than with ACPs from other systems. Similarly, the Type-II PKS ACPs form their own cluster as do both the Type-I and the Type-II FAS ACPs (Fig. 7). The determination that the SU ACPs more closely resemble Type-I FAS systems than the monofunctional FAS ACP from the same host suggests that Type-I PKSs and Type-I FASs had a common ancestor. This hypothesis is corroborated by the observation of a similar pattern when the six *eryA* KSs were compared
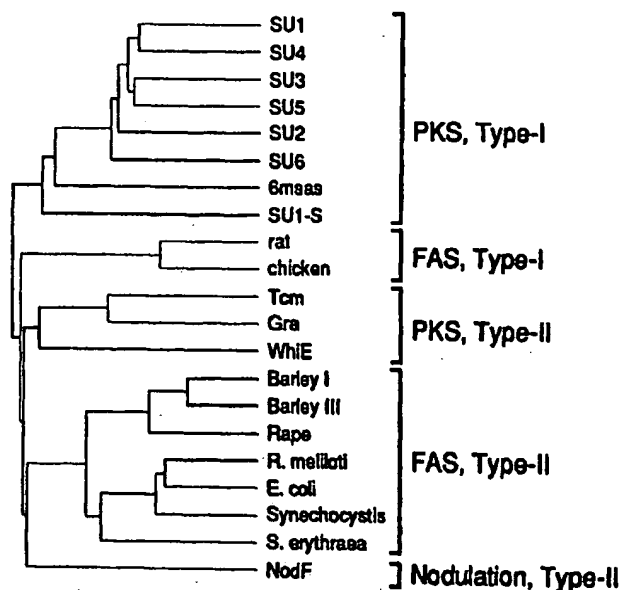
Fig. 7. Relatedness of ACPs and ACP domains. The ACP domains from multifunctional FAS and PKS systems (see Fig. 2) and the monofunctional ACPs are shown as the dendrogram obtained from PILEUP. ACP sequences: WhiE, spore-associated pigment genes from *Streptomyces coelicolor* (Davis and Chater, 1990); Barley I and III, forms I and III (Hansen, 1987); Rape, *Brassica napus* plastid seed (Safford et al., 1988); R. meliloti, *Rhizobium meliloti* constitutive ACP (Platt et al., 1990); E. coli, *E. coli* FAS ACP (Holak et al., 1988); Synechocystis, *Synechocystis* 6803 (Froehlich et al., 1990); S. erythraea, *S. erythraea* putative FAS ACP (Revill and Leadlay, 1991); NodF, *R. meliloti* nodulation-specific ACP (Debelle and Sharma, 1986).

with eight other Type-I and Type-II sequences (data not shown).

The finding of a stretch of the DH-ER interdomain region in the SU lacking these two functions is also consistent with the hypothesis that the *eryA* modules are likely to have evolved from an ancestral element (FAS- or PKS-like) which encoded the full set of activities involved in the processing of the β-carbonyl (DH, ER and KR), followed by loss of the functions not required at particular steps of 6dEB synthesis. Two modes of specialization through loss of function seem to have occurred in the *eryA* modules: selected mutations in the KR-encoding domain in module 3, and loss of the DH- and ER-encoding segments in all of the modules except module 4. Loss of function (ER) through extensive deletion may have also taken place in 6msas. It will be interesting to analyze the sequences of other PKS systems lacking KR, DH or ER domains to better understand the mode of evolution of pathways for complex polyketides.

(f) Conclusions

Our results on the extent of the various domains in the six *eryA* SU, determined solely by computer-assisted alignments, can be extended to other related systems and are substantially in agreement with those independently found by P.F. Leadlay and colleagues (personal communication) and by Witkowski et al. (1991a), who corroborated their computer analysis with limited proteolysis studies. The existence of multiple sequences with identical function in *eryA* has greatly facilitated assignments of the various domains. We have proposed a location for the DH domain and a putative active-site His for it. Type-I FAS and PKS systems also seem to share a common origin independent of their prokaryotic or eukaryotic source.

ACKNOWLEDGEMENTS

REFERENCES

Amy, C.M., Witkowski, A., Naggert, J., Williams, B., Randhawa, Z. and Smith, S.: Molecular cloning and sequencing of cDNAs encoding the entire rat fatty acid synthase. Proc. Natl. Acad. Sci. USA 86 (1989) 3114–3118.

Beck, J., Ripka, S., Siegner, A., Schiltz, E. and Schweizer, E.: The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*. Its gene structure relative to that of other polyketide synthases. Eur. J. Biochem. 192 (1990) 487–498.

Bibb, M.J., Biro, S., Motamedi, H., Collins, J.F. and Hutchinson, C.R.: Analysis of the nucleotide sequence of the *Streptomyces glaucescens tcmI* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. EMBO J. 8 (1989) 2727–2736.

Bloch, K.: β-Hydroxydecanoyl thioester dehydrase. In: Boyer, P.D. (Ed.), The Enzymes, 3rd ed., Vol. V. Academic Press, New York, 1971, pp. 441–464.

Chirala, S.S., Kuziora, M.A., Spector, D.M. and Wakil, S.J.: Complementation of mutations and nucleotide sequence of FAS1 gene encoding the β-subunit of yeast fatty acid synthase. J. Biol. Chem. 262 (1987) 4321–4340.

Cronan Jr., J.E., Li, W.-B., Coleman, R., Narasihman, M., de Mendoza, D. and Schwab, J.M.: Derived amino acid sequence and identification of the active site residues of *Escherichia coli* β-hydroxydecanoyl thioester dehydrase. J. Biol. Chem. 263 (1988) 4641–4646.

Cortes, J., Haydock, S.H., Roberts, G.A., Bevitt, D.J. and Leadlay, P.F.: An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. Nature 348 (1990) 176–178.

Davis, N.K. and Chater, K.F.: Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. Mol. Microbiol. 4 (1990) 1679–1691.

Debelle, S. and Sharma, S.B.: Nucleotide sequence of *Rhizobium meliloti* RCR2011 genes involved in host specificity of nodulation. Nucleic Acids Res. 14 (1986) 7453–7472.

Devereux, J., Haeberli, P. and Smithies, O.: A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12 (1984) 387–395.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. and Katz, L.: Modular organization of genes required for complex polyketide biosynthesis. Science 252 (1991) 675–679.

Froehlich, J.E., Poorman, R., Reardon, E., Barnum, S.R. and Jaworski, J.G.: Purification and characterization of acyl carrier protein from two cyanobacteria species. Eur. J. Biochem. 193 (1990) 817–825.

Hallam, S.E., Malpartida, F. and Hopwood, D.A.: Nucleotide sequence, transcription and deduced function of a gene involved in polyketide antibiotic synthesis in *Streptomyces coelicolor*. Gene 74 (1988) 305–320.

Hansen, L.: Three cDNA clones for barley leaf acyl carrier protein I and III. Carlsberg Res. Commun. 52 (1987) 381–392.

Hess, G.P.: Chymotrypsin – chemical properties and catalysis. In: Boyer, P.D. (Ed.), The Enzymes, 3rd ed., Vol. III. Academic Press, New York, 1971, pp. 213–248.

Holak, T.A., Nilges, M., Prestegard, J.H., Gronenborg, A.M. and Clore, G.M.: Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic resonance and the combined use of dynamical simulated annealing and distance geometry. Eur. J. Biochem. 175 (1988) 9–15.

Holzer, K., Liu, W. and Hammes, G.H.: Molecular cloning and sequencing of chicken liver fatty acid synthase cDNA. Proc. Natl. Acad. Sci. USA 86 (1989) 4387–4391.

Linial, M., Miller, K. and Scheller, R.H.: VAT-1. An abundant membrane protein from torpedo colinergic synaptic vesicles. Neuron 2 (1989) 1265–1273.

McCarthy, A.D. and Hardie, D.G.: Fatty acid synthase – an example of protein evolution by gene fusion. Trends Biochem. Sci. 9 (1984) 60–63.

Piatigorsky, J. and Wistow, G.: The recruitment of crystallins: new functions precede gene duplication. Science 252 (1991) 1078–1079.

Platt, M.W., Miller, K.J., Lane, W.S. and Kennedy, E.P.: Isolation and characterization of the constitutive acyl carrier protein from *Rhizobium meliloti*. J. Bacteriol. 172 (1990) 5440–5444.

Poulose, A.J., Rogers, L., Cheesbrough, T.M. and Kolattukudy, P.E.: Cloning and sequencing of the cDNA for S-acyl fatty acid synthase thioesterase from the uropygial gland of mallard duck. J. Biol. Chem. 260 (1985) 15953–15958.

Randhawa, Z.I. and Smith, S.: Complete amino acid sequence of the medium-chain S-acyl fatty acid synthetase thio ester hydrolase from rat mammary gland. Biochemistry 26 (1987) 1365–1373.

Revill, W.P. and Leadlay, P.F.: Cloning, characterization and high level expression in *Escherichia coli* of the *Saccharopolyspora erythraea* gene encoding an acyl carrier protein potentially involved in fatty acid biosynthesis. J. Bacteriol. 173 (1991) 4379–4385.

Rodokanaki, A., Holmes, R.K. and Borras, T.: Zeta crystallin, a novel protein from guinea pig lens is related to alcohol dehydrogenase. Gene 78 (1989) 215–224.

Safford, R., de Silva, J., Lucas, C., Windust, J.H.C., Shedden, J., James, C.M., Sidebottom, C.M., Slabas, A.R., Tombs, M.P. and Huges, S.G.: Molecular cloning and sequence analysis of complementary DNA encoding rat mammary gland medium-chain S-acyl fatty acid synthetase thio ester hydrolase. Biochemistry 26 (1987) 1358–1364.

Safford, R., Windust, J.H.C., Lucas, C., de Silva, J., James, C.M., Hellyer, A., Smith, C.G., Slabas, A.R. and Huges, S.G.: Plastid-localised seed acyl-carrier protein of *Brassica napus* is encoded by a distinct, nuclear multigene family. Eur. J. Biochem. 174 (1988) 287–295.

Schweizer, M., Roberts, L.M., Holtke, H.-J., Takabayashi, K., Hollerer, E., Hoffmann, B., Muller, G., Kottig, H. and Schweizer, E.: The pentafunctional FAS1 gene of yeast: its nucleotide sequence and order of catalytic domains. Mol. Gen. Genet. 203 (1986) 479–486.

Sherman, D.H., Malpartida, F., Bibb, M.J., Kieser, H.H., Bibb, M.J. and Hopwood, D.A.: Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tu22. EMBO J. 8 (1989) 2717–2725.

Tsukamoto, Y. and Wakil, S.J.: Isolation and mapping of the β-hydroxyacyl dehydratase activity of chicken liver fatty acid synthase. J. Biol. Chem. 263 (1988) 16225–16229.

Wakil, S.J.: Fatty acid synthase, a proficient multifunctional enzyme. Biochemistry 28 (1989) 4523–4530.

Weber, J.M., Leung, J.O., Swanson, S.J., Idler, K.B. and McAlpine, J.B.: An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. Science 252 (1991) 114–117.

Witkowski, A., Rangan, V.S., Randhawa, Z.I., Amy, C.M. and Smith, S.: Structural organization of the multifunctional animal fatty-acid synthase. Eur. J. Biochem. 198 (1991a) 571–579.

Witkowski, A., Naggert, J., Wessa, B. and Smith, S.: A catalytic role for histidine-237 in rat mammary gland thioesterase II. J. Biol. Chem. 266 (1991b) 18514–18519.

Wong, A., Kean, L. and Maurer, R.: Sequence of the *dnaB* gene of *Salmonella typhimurium*. J. Bacteriol. 170 (1988) 2668–2675.

Yuan, Z., Liu, W. and Hammes, G.H.: Molecular cloning and sequencing of DNA complementary to chicken liver fatty acid synthase mRNA. Proc. Natl. Acad. Sci. USA 85 (1988) 6328–6331.

Insertional Inactivation of Methylmalonyl Coenzyme A (CoA) Mutase and Isobutyryl-CoA Mutase Genes in *Streptomyces cinnamonensis*: Influence on Polyketide Antibiotic Biosynthesis

JAN W. VRIJBLOED, KATJA ZERBE-BURKHARDT, ANANDA RATNATILLEKE,
ANDREAS GRUBELNIK-LEISER, AND JOHN A. ROBINSON*

Department of Chemistry, University of Zurich, CH-8057 Zurich, Switzerland

The coenzyme B_{12}-dependent isobutyryl coenzyme A (CoA) mutase (ICM) and methylmalonyl-CoA mutase (MCM) catalyze the isomerization of *n*-butyryl-CoA to isobutyryl-CoA and of methylmalonyl-CoA to succinyl-CoA, respectively. The influence that both mutases have on the conversion of *n*- and isobutyryl-CoA to methylmalonyl-CoA and the use of the latter in polyketide biosynthesis have been investigated with the polyether antibiotic (monensin) producer *Streptomyces cinnamonensis*. Mutants prepared by inserting a hygromycin resistance gene (*hygB*) into either *icmA* or *mutB*, encoding the large subunits of ICM and MCM, respectively, have been characterized. The *icmA::hygB* mutant was unable to grow on valine or isobutyrate as the sole carbon source but grew normally on butyrate, indicating a key role for ICM in valine and isobutyrate metabolism in minimal medium. The *mutB::hygB* mutant was unable to grow on propionate and grew only weakly on butyrate and isobutyrate as sole carbon sources. ^{13}C-labeling experiments show that in both mutants butyrate and acetoacetate may be incorporated into the propionate units in monensin A without cleavage to acetate units. Hence, *n*-butyryl-CoA may be converted into methylmalonyl-CoA through a carbon skeleton rearrangement for which neither ICM nor MCM alone is essential.

Streptomycetes produce a wide variety of commercially important polyketide secondary metabolites, including the well-known macrolide and polyether antibiotics. The assembly of these antibiotics, catalyzed by the large family of polyketide synthases, requires malonyl coenzyme A (CoA), methylmalonyl-CoA, and ethylmalonyl-CoA as extender units, which are the precursors of the acetate-, propionate-, and butyrate-derived units, respectively, in the polyketide backbone (8, 11, 12).

Fatty acid and amino acid catabolisms are known to play an important role in supplying building blocks for polyketide biosynthesis. *n*-Butyryl-CoA and isobutyryl-CoA are known intermediates in fatty acid and valine catabolism, and both may be converted intact into methylmalonyl-CoA and ethylmalonyl-CoA in streptomycetes (Fig. 1). This is evidenced by the efficient incorporation of ^{13}C-labeled *n*- and isobutyrate into both the propionate- and butyrate-derived units in several polyketide antibiotics, including monensin A (20, 22), tylosin (18), and narasin (6). The coenzyme B_{12}-dependent isobutyryl-CoA mutase (ICM) appears to play a key role in these processes, by catalyzing the interconversion of *n*- and isobutyryl-CoA. However, the preferred route(s) for the conversion of *n*- and isobutyryl-CoA into methylmalonyl-CoA is less clearly defined. One possibility, consistent with results of ^{13}C-labeling studies, is the direct oxidation of isobutyryl-CoA, in several enzymatic steps, perhaps via methacrylyl-CoA and 3-hydroxybutyryl-CoA, to give methylmalonyl-CoA (path A in Fig. 1), although this route is different from the well-known pathway of valine catabolism in pseudomonads and mammals that leads via methylmalonic acid semialdehyde to propionyl-CoA (10, 15, 32) (Fig. 1). Another possibility is that *n*-butyryl-CoA might be oxidized at C-4 (ω oxidation) over several steps without

fragmentation, to give succinyl-CoA, which may then be converted to methylmalonyl-CoA by methylmalonyl-CoA mutase (MCM) (path B in Fig. 1). Known routes to methylmalonyl-CoA presently include the carboxylation of propionyl-CoA, catalyzed by propionyl-CoA carboxylase, and the isomerization of succinyl-CoA, catalyzed by the coenzyme B_{12}-dependent MCM. Ethylmalonyl-CoA appears to be derived solely by carboxylation of *n*-butyryl-CoA.

To further explore the roles of MCM and ICM in the production of methylmalonyl-CoA, we report here the properties of *Streptomyces cinnamonensis* mutants in which either MCM or ICM has been inactivated by gene disruption. The genes encoding both enzymes have been cloned and sequenced from *S. cinnamonensis*, the producer of the polyether antibiotic monensin A (Fig. 1) (2, 33). MCM consists of two subunits, MutA and MutB, of 65 and 79 kDa, respectively, which show high sequence similarities to MCMs from other bacteria and mammals. So far, ICM has been isolated only from *S. cinnamonensis*, where it was found to comprise a large subunit of 65 kDa (IcmA) and a small subunit of ca. 17 kDa (IcmB) (33). The cloning and insertional inactivation of *icmA* from *S. cinnamonensis* were reported earlier (33). In this work, a mutant was prepared by inserting a hygromycin resistance gene (*hygB*) into *mutB*. The growth characteristics of both *icmA::hygB* and *mutB::hygB* mutants and the pattern of incorporation of ^{13}C-labeled precursors into the polyketide monensin A in each are reported below. The results indicate a key role for ICM in valine and isobutyrate metabolism under normal growth conditions.

MATERIALS AND METHODS

---

* Corresponding author. Mailing address: Institute of Organic Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. Phone: (0041)-1-635-4242. Fax: (0041)-1-635-6812. E-mail: robinson@oci.unizh.ch.
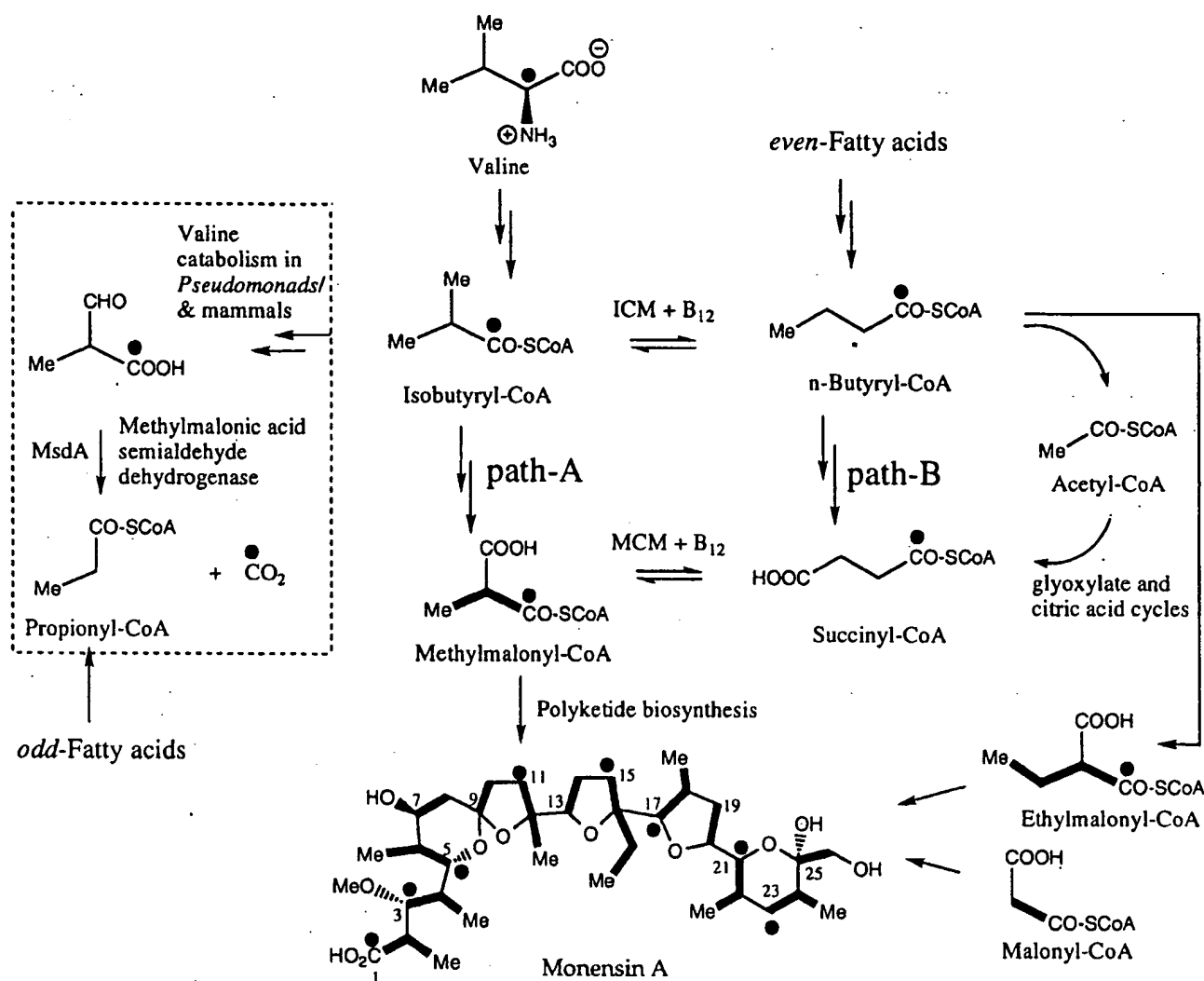
FIG. 1. Pathways of valine catabolism and the reactions catalyzed by MCM and ICM.

Insertional inactivation of *mutB*. In order to disrupt *mutB* in *S. cinnamonensis*, a 2.55-kb *Bam*HI fragment (2) containing almost the entire *mutB* gene was cloned into the *Bam*HI site of pUC18 to generate pOCI403. Subsequently, a hygromycin resistance gene (*hygB*) was cloned as a *Bam*HI/*Bgl*II fragment into the unique *Bgl*II site of pOCI403 to generate pOCI352. The *Bam*HI fragment of pOCI352 was isolated and cloned into the unique *Bam*HI cloning site of pGM160. The resulting plasmid, pOCI353, after passage through *Escherichia coli* ET12567 and *Streptomyces lividans* TK64, was used to transform *S. cinnamonensis* A3823.5. Selection for the disrupted *mutB* gene was performed as previously described (33).

Incorporation of ^{13}C-labeled precursors into monensin A. Singly labeled precursors (100 mg per culture) were fed batchwise in equal portions.after 24, 36, and 48 h to two 60-ml cultures in a complex liquid medium (22, 28). Subsequently, monensin A was purified from the culture broth as in earlier work (22) and examined by inverse-gated ^{13}C{^{1}H} nuclear magnetic resonance (NMR) spectroscopy (20-s relaxation delay, 30° pulse) on a Bruker AMX600 spectrometer. Enrichment levels were determined by signal integration. The ethyl [1,3-^{13}C$_2$]acetoacetate (100 mg) was mixed with unlabeled ethyl acetoacetate (200 mg) and then batch fed to two liquid cultures (150 mg per culture) in three equal portions after 48, 60, and 72 h of growth. Typically, 30 to 70 mg of pure monensin A was obtained from each experiment.

RESULTS

Insertional inactivation of *S. cinnamonensis* *mutB*. Disruption of *mutB* was achieved by cloning a hygromycin resistance gene, *hygB*, into the unique *Bgl*II site of *mutB* (2). A restriction fragment containing the disrupted *mutB* and *hygB* genes was then cloned into pGM160, which carries also the thiostrepton resistance gene. The resulting plasmid, pOCI353, was used to transform protoplasts of *S. cinnamonensis* A3823.5. Isolation of the desired mutant was performed by selecting for Ts^s and Hg^r colonies at 39°C, as in earlier work (33). A Southern blot hybridization analysis with total DNA isolated from one clone (*mutB::hygB*) confirmed that the *mutB* gene had been disrupted, consistent with a double-crossover event (data not shown). The *mutB::hygB* mutant was devoid of MCM activity in cell extracts.

Phenotype analysis of the *S. cinnamonensis* mutant strains. The growth characteristics of wild-type (wt) *S. cinnamonensis* and the *mutB::hygB* and *icmA::hygB* mutants on solid minimal medium containing a single carbon source were determined, and the results are summarized in Table 2. When grown in a

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Description	Reference or source
Streptomyces spp.		
S. cinnamonensis A3823.5	Monensin-overproducing strain	Eli Lilly & Co.
S. lividans TK64	*pro-2 str-6*	9
E. coli ET12567	*hsdS dam-3 dcm-6*	14
Plasmids		
pUC18		17
pGM160	Temperature-sensitive plasmid	16
pIJ963	Contains the *hygB* gene	13
pOCI403	Contains 2.3-kb *Bam*HI fragment of *mutB* in pUC18	This work
pOCI352	Hygromycin resistance gene cloned as *Bam*HI-*Bgl*II fragment in the *Bgl*II site of pOCI403	This work
pOCI353	Contains the disrupted *mutB* gene cloned as *Bam*HI fragment in pGM160	This work

complex oil-based medium (medium 5, described in reference 28), the wt and the two mutants grew equally well and produced comparable levels of monensin A. However, during growth in this medium addition of isobutyrate or valine at a ca. 5 mM concentration to the *icmA::hygB* mutant caused cessation of further growth and monensin production. In a chemically defined medium (4) containing mainly glucose, tyrosine, and valine as C and N sources, the growth of the *mutB::hygB* mutant was comparable to that of the wt, whereas the growth of the *icmA::hygB* mutant was weak.

Incorporation of ^{13}C-labeled precursors into monensin A. The ability of *n*-butyrate and isobutyrate to serve as precursors for monensin A biosynthesis in cultures of the mutant strains *mutB::hygB* and *icmA::hygB* was tested by adding *n*-[1-^{13}C]butyrate and [1-^{13}C]isobutyrate to shake cultures of each, as well as to wt *S. cinnamonensis*. Monensin A was subsequently isolated and examined by ^{13}C{^{1}H} NMR spectroscopy. The observed enrichments are summarized in Table 3. Only trace amounts of monensin A could be isolated from the *icmA::hygB* strain when it was fed with [1-^{13}C]isobutyrate. The addition of isobutyrate (or valine) to this mutant strongly inhibited further growth and monensin production. The same effect was not observed upon feeding the mutant with *n*-[1-^{13}C]butyrate. The *mutB::hygB* strain showed relatively high incorporations of ^{13}C

in the propionate and butyrate units of monensin A upon being fed with both labeled butyrate and isobutyrate (Table 3).

Ethyl [1,3-^{13}C$_2$]acetoacetate was fed to the wt and both *S. cinnamonensis* mutants, the monensin A was subsequently isolated from each culture, and ^{13}C{^{1}H} NMR spectra were recorded. For monensin A from both mutants, highly enriched doublet signals were observed surrounding the natural abundance singlets for all positions derived from C-1 and C-2 of a propionate building block, i.e., in monensin A, C-1–C-2, C-3–C-4, C-5–C-6, C-11–C-12, C-17–C-18, C-21–C-22, and C-23–C-24 (see Fig. 2 for examples). This implies the formation of [1,2-^{13}C$_2$]methylmalonyl-CoA from [1,3-^{13}C$_2$]acetoacetyl-CoA in vivo. Also, the signals for C-15 and C-32, corresponding to the C-1 and C-3 positions in the butyrate building block, were strongly enriched (Fig. 2).

DISCUSSION

ICM is an MCM-like enzyme from *S. cinnamonensis* comprising a large subunit (IcmA) of 566 residues (33) and a small

TABLE 2. Growth of *S. cinnamonensis* strains on solid minimal medium with various added carbon sources[a]

Carbon Source	wt	*mutB::hygB*	*icmA::hygB*
Glucose	+++	+++	+++
Succinate	+++	+++	+++
Valine	++	++	−
Acetate	++	++	++
Propionate	+++	−	+++
Butyrate	+++	+	+++
Valerate	+++	−	+++
Caproate	++	+	++
Isobutyrate	++	+	−
Isocaproic acid	++	+	−
Crotonic acid	++	+	++
Ethyl acetoacetate	++	+	++
rac-3-Hydroxybutyrate	+++	++	+++
None	−	−	−

[a] +++, strong, rapid, dense mycelial growth; ++, intermediate growth; +, slow, weak, but still clear growth; −, almost no mycelial growth, corresponding to growth without a carbon source.

TABLE 3. Enrichments observed in monensin A after incorporation of *n*-[1-^{13}C]butyrate and [1-^{13}C]isobutyrate in the wt and in *mutB::hygB* and *icmA::hygB* mutants of *S. cinnamonensis*

| Monensin carbon	*n*-[1-^{13}C]butyrate			[1-^{13}C]isobutyrate		
	wt	*mutB::hyg*	*icmA::hyg*	wt	*mutB::hyg*	*icmA::hyg*
Acetate unit						
C-7	1.4	1.5	1.3	1.7	1.6	ND[a]
C-9	1.5	1.6	1.6	2.1	1.8	
C-13	1.5	1.3	1.4	2.8	1.7	
C-19	1.8	1.8	1.5	3.4	2.2	
C-25	1.6	1.6	1.3	2.1	1.8	
Propionate unit						
C-1	5.0	8.8	3.1	6.9	9.2	ND
C-3	5.1	8.8	3.0	7.7	9.3	
C-5	5.2	9.2	3.1	7.2	9.3	
C-11	4.6	8.8	3.1	8.2	8.6	
C-17	4.7	8.7	3.1	6.8	9.3	
C-21	5.3	9.5	3.2	7.1	9.4	
C-23	4.9	8.9	3.1	7.1	9.0	
Butyrate unit						
C-15	16.8	12.3	8.9	17.0	12.4	ND

[a] ND, not determined (see text).
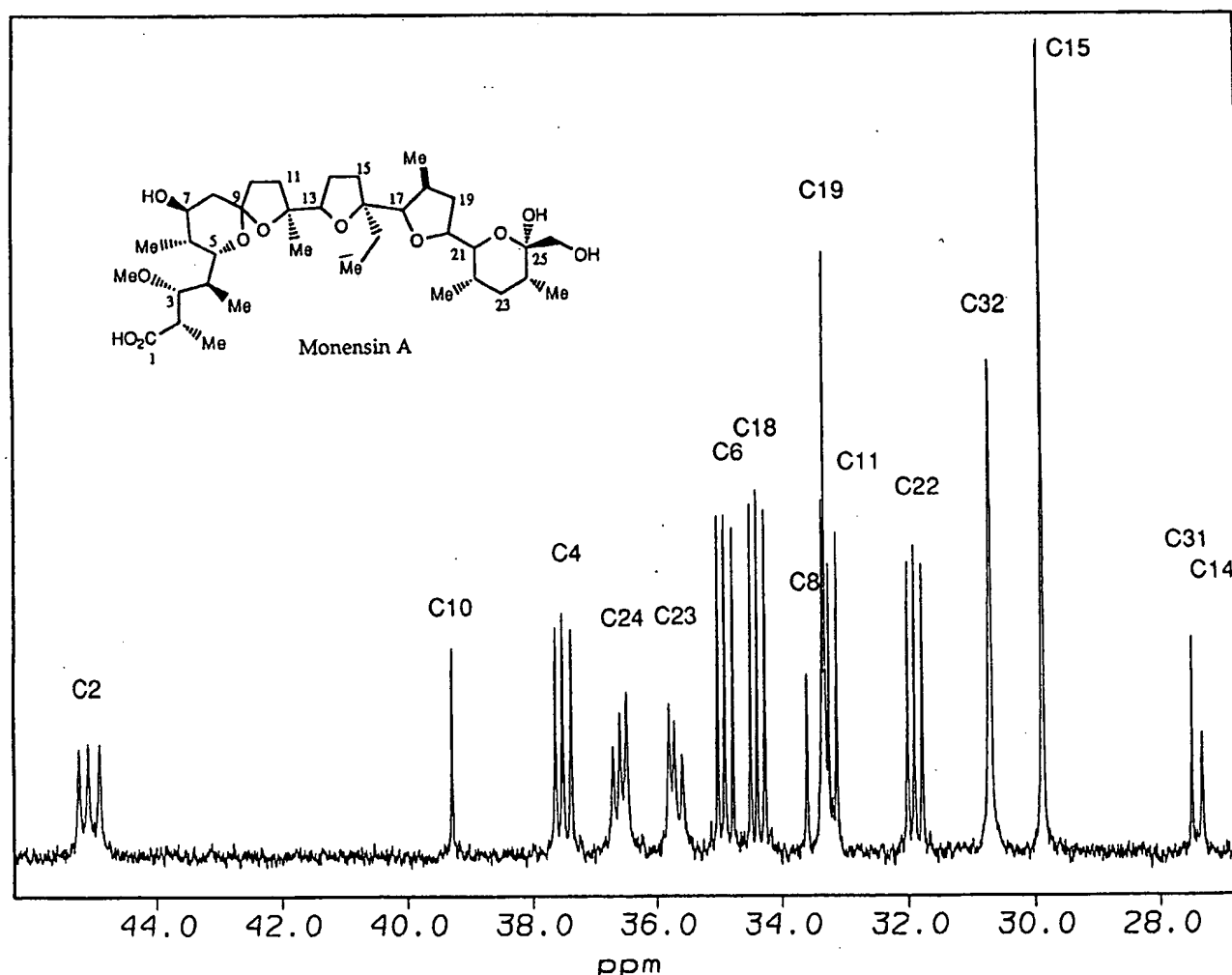
FIG. 2. A portion of the $^{13}C\{^1H\}$ NMR spectrum of monensin A isolated after feeding $[1,3-^{13}C_2]$ethyl acetoacetate to a *mutB::hygB* mutant of *S. cinnamonensis*. The strongly enriched signals from the *n*-butyrate unit in the backbone of monensin (at positions 15 and 32) are shown, along with the enriched doublets (for C-2, C-4, C-24, C-23, C-6, C-18, C-11, and C-22) surrounding the natural abundance signals for positions derived from propionate building blocks.

subunit (IcmB) of 136 residues (21). The MCM from this organism is a heterodimer with a large subunit (MutB) of 733 residues and a small subunit (MutA) of 616 residues (2). In this work, MCM was inactivated in *S. cinnamonensis* by inserting a *hygB* gene into *mutB*, to produce an *mutB::hygB* mutant. The cloning and insertional inactivation of *icmA* from *S. cinnamonensis* were reported earlier (33).

The effect of the *icmA* and *mutB* mutations on the growth of *S. cinnamonensis* was studied on solid minimal medium with various carbon sources. Both mutants show comparable growth on control plates containing glucose and succinate (Table 2). As with a previously reported knockout of MCM in *Nocardia corallina* (30), the *S. cinnamonensis mutB::hygB* mutant was unable to grow on propionate and valerate, since the conversion of methylmalonyl-CoA to succinyl-CoA was blocked (Fig. 1). The growth of this mutant was also strongly retarded on *n*- and isobutyrate, *n*- and isocaproate, and acetoacetate and crotonic acid as sole carbon sources.

The *icmA::hygB* mutant was unable to grow on valine, isobutyrate, or isocaproate, although it grew normally on acetate, butyrate, and valerate (Table 2). Moreover, addition of isobutyrate during growth in rich medium in liquid culture caused

cessation of further growth and monensin production (see labeling studies in Materials and Methods and Results). The ability of both *mutB::hygB* and *icmA::hygB* mutants to grow normally on acetate as sole carbon source is inconsistent with the operation of a recently proposed (7) novel anaplerotic pathway that is an alternative to the glyoxylate cycle, for utilization of C_2 and C_4 fatty acids in streptomycetes.

Valine catabolism in pseudomonads and mammals (10, 15, 32) proceeds via isobutyryl-CoA, 3-hydroxybutyric acid, and methylmalonic acid semialdehyde, to propionyl-CoA (indicated in the dotted box in Fig. 1). The gene encoding methylmalonic acid semialdehyde dehydrogenase was cloned recently from *Streptomyces coelicolor* (34). An *msdA::hygA* mutant lost methylmalonic acid semialdehyde dehydrogenase activity as well as the ability to grow in a minimal medium with valine or isobutyrate as the sole carbon source. Nevertheless, labeling studies of many (but not all [19]) polyketide antibiotic-producing streptomycetes (3, 18, 20, 22, 24) have shown that $[2-^{13}C]$valine and $[1-^{13}C]$isobutyrate label C-1 of each propionate-derived subunit in the polyketide backbone (as indicated for monensin A in Fig. 1), rather than being lost as carbon dioxide. The results of these labeling experiments are incon-

sistent with a major flux through the *Pseudomonas*-mammalian pathway of valine catabolism during polyketide biosynthesis in the stationary phase in these streptomycetes, including *S. cinnamonensis*. Moreover, since the *S. cinnamonensis icmA::hygB* mutant was unable to grow on valine or isobutyrate, this *Pseudomonas*-mammalian pathway of valine catabolism is either absent or poorly expressed in this mutant during growth on minimal medium.

The conversion of isobutyryl-CoA to methylmalonyl-CoA by path A (Fig. 1) has been frequently discussed in studies of antibiotic-producing streptomycetes (29), including *S. cinnamonensis* (22). If path A is operative, however, this should provide a route for the utilization of valine and isobutyrate in the *icmA::hygB* mutant, and yet the mutant could not grow in minimal medium with isobutyrate or valine as sole carbon source (Table 2). It seems possible that, during antibiotic production in the stationary phase in a rich complex medium (28), a different pattern of gene expression and metabolite utilization may occur, compared to the situation on minimal agar. The inability of the *icmA::hygB* mutant to grow on isobutyrate and valine in minimal medium, therefore, may not rule out a role for path A (Fig. 1) in complex medium during the stationary phase in *S. cinnamonensis*.

An alternative route to methylmalonyl-CoA via succinyl-CoA might conceivably involve the conversion of isobutyryl-CoA to *n*-butyryl-CoA, with ω oxidation of the latter to succinyl-CoA (path B in Fig. 1). The reversible conversion of crotonyl-CoA to 4-hydroxybutyryl-CoA catalyzed by an oxygen-sensitive, flavin-containing enzyme is known for clostridia (23, 26). but it is so far unclear whether a similar enzyme occurs in streptomycetes, and no other route for the ω functionalization of *n*-butyryl-CoA is presently known. In addition, it was shown earlier (27) that trideuteriolabeled acetate can be efficiently incorporated into all the propionate units in monensin A without loss of deuterium. This observation rules out a major flux from *n*-butyryl-CoA to methylmalonyl-CoA via succinyl-CoA but is consistent with path A (Fig. 1). It should be noted, however, that similar experiments carried out with *Streptomyces longisporoflavus*, the producer of the polyether ICI139603 (also called tetronasin), showed that the C-methyl groups retain at most two deuterium atoms from acetate (5), as would be expected for an incorporation of the label via succinate and MCM (Fig. 1).

To determine how *n*- and isobutyrate are utilized for polyketide biosynthesis in the *icmA* and *mutB* mutant strains, [13C]-labeled precursors were fed to each, and the pattern of incorporation into monensin A was determined by 13C NMR. An efficient incorporation of *n*-[1-13C]butyrate into the propionate-derived units in the backbone of monensin A was observed in the wt and was higher in the *mutB* mutant but slightly lower in the *icmA* mutant (Table 3). The enrichments in acetate-derived units in monensin A were low, indicating a minimal breakdown of [13C]butyrate to [13C]acetate prior to incorporation. These results are consistent with a major flux through path A (Fig. 1) but show also that label is still incorporated at a lower efficiency, by an alternative route, when ICM is inactivated.

The efficient incorporation of [1-13C]isobutyrate into the propionate units in monensin A in the *mutB::hygB* strain shows that the conversion of isobutyryl-CoA to methylmalonyl-CoA does not require MCM. Unfortunately, it was not possible to incorporate [1-13C]isobutyrate into monensin in the *icmA* mutant, because addition of isobutyrate (or valine) to the culture arrested further growth and monensin production.

Finally, a labeling experiment with ethyl [1,3-13C_2]acetoacetate was carried out to establish whether the incorporation of

a C_4 precursor into the propionate units is possible without prior cleavage to acetate units. Most likely, the precursor is hydrolyzed by endogenous lipases or esterases, is activated in the cytoplasm to [1,3-13C_2]acetoacetyl-CoA, and then is converted to *n*-butyryl-CoA by steps similar to those involved in fatty acid metabolism (31) or polyhydroxybutyrate metabolism (1). The monensin A isolated from both mutants and the wt showed an essentially identical pattern of one-bond 13C-13C couplings (Fig. 2), within each propionate unit, consistent with the formation and incorporation of [1,2-13C_2]methylmalonyl-CoA without fragmentation of the precursor to acetate units. Hence, [1,3-13C_2]acetoacetyl-CoA can be converted efficiently into [1,2-13C_2]methylmalonyl-CoA in the absence of either ICM or MCM. Presently, it is uncertain how butyryl-CoA is converted intact into methylmalonyl-CoA when ICM is absent, although we disfavor path B in Fig. 1 for reasons discussed above.

At the present time, we are unable to identify one unifying metabolic plan that fully explains all the data discussed above. A scenario, however, in which a still unrecognized route allows the conversion of *n*-butyryl-CoA (or acetoacetyl-CoA) into methylmalonyl-CoA in streptomycetes may be considered. In this respect, it is interesting to note that a third MCM-like protein of unknown function called MEA has been described recently (7). MEA has been implicated in the assimilation of C_2 compounds in *Streptomyces collinus* and in methanol and ethanol utilization in *Methylobacterium extorquens* (25). Further clarification of C_2 and C_4 metabolism in streptomycetes may benefit from the elucidation of the function of this MEA protein.

REFERENCES

1. **Anderson, A. J., and E. A. Dawes.** 1990. Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. Microbiol. Rev. 54:450–472.
2. **Birch, A., A. Leiser, and J. A. Robinson.** 1993. Cloning, sequencing, and expression of the gene encoding methylmalonyl-coenzyme A mutase from *Streptomyces cinnamonensis.* J. Bacteriol. 175:3511–3519.
3. **Clark, C. A., and J. A. Robinson.** 1985. Biosynthesis of nonactin. The role of acetoacetyl-CoA in the formation of nonactic acid. J. Chem. Soc. Chem. Commun., p. 1568–1569.
4. **Day, L. E., J. W. Chamberlin, E. Z. Gordee, S. Chen, M. Gorman, R. L. Hamill, T. Ness, R. E. Weeks, and R. Stroshane.** 1973. Biosynthesis of monensin. Antimicrob. Agents Chemother. 4:410–414.
5. **Doddrell, D. M., E. D. Laue, F. J. Leeper, J. Staunton, A. Davies, A. B. Davies, and G. A. F. Ritchie.** 1984. Biosynthesis of the polyether antibiotic ICI139603 in *Streptomyces longisporoflavus*: investigation of deuterium retention after incorporation of CD_3^13COOH, 13CD_3COOH and CH_3CD_2^13COOH using 2H NMR and edited 13C NMR spectra. J. Chem. Soc. Chem. Commun., p. 1302–1304.
6. **Dorman, D. E., J. W. Paschal, W. M. Nakatsukasa, L. L. Huckstep, and N. Neuss.** 1976. The use of 13C-NMR spectroscopy in biosynthetic studies. II. Biosynthesis of narasin, a new polyether ionophore from fermentation of *Streptomyces aureofaciens*. Helv. Chim. Acta 59:2625–2634.
7. **Han, L., and K. A. Reynolds.** 1997. A novel alternate anaplerotic pathway to the glyoxylate cycle in streptomycetes. J. Bacteriol. 179:5157–5164.
8. **Hopwood, D. A.** 1997. Genetic contributions to understanding polyketide synthases. Chem. Rev. 97:2465–2497.
9. **Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. J. Bruton, H. M. Kieser, D. J. Lydiate, C. P. Smith, J. M. Ward, and H. Schrempf.** 1985. Genetic manipulation of *Streptomyces*: a laboratory manual. John Innes Foundation, Norwich, United Kingdom.
10. **Kamoun, P.** 1992. Valine is a precursor of propionyl-CoA. Trends Biochem. Sci. 17:175.
11. **Katz, L.** 1997. Manipulation of modular polyketide syntheses. Chem. Rev. 97:2557–2575.
12. **Leadlay, P. F.** 1997. Combinatorial approaches to polyketide biosynthesis. Curr. Opin. Chem. Biol. 1:162–168.

13. Lydiate, D. J., A. M. Ashby, D. J. Henderson, H. M. Kieser, and D. A. Hopwood. 1989. Physical and genetic characterization of chromosomal copies of the *Streptomyces coelicolor* mini-circle. J. Gen. Microbiol. 135:941–955.

14. MacNeil, D. J. 1988. Characterization of a unique methyl-specific restriction system in *Streptomyces avertimilis*. J. Bacteriol. 170:5607–5612.

15. Massey, L. K., J. R. Sokatch, and R. S. Conrad. 1976. Branched-chain amino acid catabolism in bacteria. Bacteriol. Rev. 40:42–54.

16. Muth, G., B. Nussbaumer, W. Wohlleben, and A. Pühler. 1989. A vector system with temperature sensitive replication for gene disruption and mutational cloning in Streptomyces. Mol. Gen. Genet. 219:341–348.

17. Norrander, J., T. Kempe, and J. Messing. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101–106.

18. Omura, S., K. Tsuzuki, Y. Tanaka, H. Sakakibara, M. Aizawa, and G. Lukacs. 1983. Valine as a precursor of n-butyrate unit in the biosynthesis of macrolide aglycone. J. Antibiot. 36:614–616.

19. Ono, M., H. Mishima, Y. Takiguchi, and M. Terao. 1983. Milbemycins, a new family of macrolide antibiotics. Studies on the biosynthesis of milbemycins a_2, a_4 and D using ^{13}C labeled precursors. J. Antibiot. 36:991–1000.

20. Pospíšil, S., P. Sedmera, M. Havránek, V. Krumphanzl, and Z. Vaněk. 1983. Biosynthesis of monensins A and B. J. Antibiot. 36:617–619.

21. Ratnatilleke, A., W. J. Vrijbloed, and J. A. Robinson. Cloning of the coenzyme B_{12}-binding domain of isobutyryl-CoA mutase from *Streptomyces cinnamonensis*, reconstitution of mutase activity and characterization of the recombinant enzyme produced in *Escherichia coli*. Submitted for publication.

22. Reynolds, K. A., D. O'Hagan, D. Gani, and J. A. Robinson. 1988. Butyrate metabolism in streptomycetes. Characterization of an intramolecular interchange rearrangement linking isobutyrate and butyrate in *Streptomyces cinnamonensis*. J. Chem. Soc. Perkin Trans. 1, p. 3195–3207.

23. Scherf, U., and W. Buckel. 1993. Purification and properties of an iron-sulfur and FAD-containing 4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA Δ^3-Δ^2-isomerase from *Clostridium aminobutyricum*. Eur. J. Biochem. 275:421–429.

24. Sherman, M. M., S. Yue, and C. R. Hutchinson. 1986. Biosynthesis of lasalocid A. Metabolic interrelationships of carboxylic acid precursors and polyether antibiotics. J. Antibiot. 39:1135–1143.

25. Smith, L. M., W. G. Meijer, L. Dijkhuizen, and P. M. Goodwin. 1996. A protein having similarity with methylmalonyl-CoA mutase is required for the assimilation of methanol and ethanol by *Methylobacterium extorquens* AM1. Microbiology 142:675–684.

26. Söhling, B., and G. Gottschalk. 1996. Molecular analysis of the anaerobic succinate degradation pathway in *Clostridium kluyveri*. J. Bacteriol. 178:871–880.

27. Sood, G. R., D. M. Ashworth, A. A. Ajaz, and J. A. Robinson. 1988. Biosynthesis of the polyether antibiotic monensin A. Results from the incorporations of labelled acetate and propionate as probes of the carbon chain assembly processes. J. Chem. Soc. Perkin Trans. 1, p. 3183–3193.

28. Stark, W. M., N. G. Knox, and J. E. Westhead. 1968. Monensin, a new biologically active compound. II. Fermentation studies, p. 353–358. Antimicrob. Agents Chemother. 1967.

29. Tang, L., Y.-X. Zhang, and C. R. Hutchinson. 1994. The genetic basis of precursor supply for the biosynthesis of macrolide and polyether antibiotics, p. 105–116. *In* R. K. Bajpai and A. Prokop (ed.), Recombinant DNA technology II. The New York Academy of Sciences, New York, N.Y.

30. Valentin, H. E., and D. Dennis. 1996. Metabolic pathway for poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate) formation in *Nocardia corallina*: inactivation of *mutB* by chromosomal integration of a kanamycin resistance gene. Appl. Environ. Microbiol. 62:372–379.

31. Wallace, K. K., Z.-Y. Bao, H. Dai, R. Digate, G. Schuler, M. K. Speedie, and K. A. Reynolds. 1995. Purification of crotonyl-CoA reductase from *Streptomyces collinus* and cloning, sequencing and expression of the corresponding gene in *Escherichia coli*. Eur. J. Biochem. 233:954–962.

32. Wolf, D. A., and H. A. Akers. 1986. Uncertainties remain in the catabolism of valine. Trends Biochem. Sci. 11:390–392.

33. Zerbe Burkhardt, K., A. Ratnatilleke, N. Philippon, A. Birch, A. Leiser, J. W. Vrijbloed, D. Hess, P. Hunziker, and J. A. Robinson. 1998. Cloning, sequencing, expression, and insertional inactivation of the gene for the large subunit of the coenzyme B_{12}-dependent isobutyryl-CoA mutase from *Streptomyces cinnamonensis*. J. Biol. Chem. 273:6508–6517.

34. Zhang, Y. X., L. Tang, and C. R. Hutchinson. 1996. Cloning and characterization of a gene (*msdA*) encoding methylmalonic acid semialdehyde dehydrogenase from *Streptomyces coelicolor*. J. Bacteriol. 178:490–495.

Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering

D. L. Stassi†‡, S. J. Kakavas†, K. A. Reynolds§, G. Gunawardana†, S. Swanson†, D. Zeidner†, M. Jackson†, H. Liu§, A. Buko†, and L. Katz†

†Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, IL 60064; and §Institute of Structural Biology and Drug Discovery, Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23219

ABSTRACT   A previously unknown chemical structure, 6-desmethyl-6-ethylerythromycin A (6-ethylErA), was produced through directed genetic manipulation of the erythromycin (Er)-producing organism *Saccharopolyspora erythraea*. In an attempt to replace the methyl side chain at the C-6 position of the Er polyketide backbone with an ethyl moiety, the methylmalonate-specific acyltransferase (AT) domain of the Er polyketide synthase was replaced with an ethylmalonate-specific AT domain from the polyketide synthase involved in the synthesis of the 16-member macrolide niddamycin. The genetically altered strain was found to produce ErA, however, and not the ethyl-substituted derivative. When the strain was provided with precursors of ethylmalonate, a small quantity of a macrolide with the mass of 6-ethylErA was produced in addition to ErA. Because substrate for the heterologous AT seemed to be limiting, crotonyl-CoA reductase, a primary metabolic enzyme involved in butyryl-CoA production in streptomycetes, was expressed in the strain. The primary macrolide produced by the reengineered strain was 6-ethylErA.

Erythromycin (Er) is a broad-spectrum macrolide antibiotic produced by *Saccharopolyspora erythraea*. The backbone of the molecule is a 14-member macrocyclic ring (Fig. 1) that is produced through the sequential condensation of one molecule of propionyl-CoA and six molecules of methylmalonyl-CoA by a modular polyketide synthase (PKS). This enzyme complex comprises three large multifunctional polypeptides, DEBS1, DEBS2, and DEBS3, each of which contains two modules, and, in the case of DEBS1, a loading domain to initiate synthesis of the polyketide chain (1, 2). Each module contains the enzymatic activities necessary for one condensation and subsequent reduction of the extender to the growing chain. Selection of the appropriate extender unit is accomplished by the acyltransferase (AT) domain present in each module (3).

Recently, genetic engineering has joined synthetic chemistry in the production of novel macrolide structures. Hybrid PKSs have been constructed through the replacement of AT domains with those that specify different starter or extender units (3–6). These manipulations have resulted in the production of Er derivatives lacking methyl groups at C-10 and C-12 (5) and those incorporating the branched chain starter units of avermectin biosynthesis (6). To date, however, there have been no manipulations that allow the methyl side chains of Er to be replaced by ethyl groups. This type of modification would be useful in expanding the structural diversity of hybrid polyketides produced by combinatorial biosynthesis.

Niddamycin (Nd) is a 16-member macrolide with an ethyl side chain at C-6 (Fig. 1). The module 5 AT, which is believed to be responsible for incorporation of the ethyl side chain into the polyketide backbone, was identified by sequence analysis of the Nd PKS genes (7). In this paper we describe the construction of a hybrid PKS through the replacement of a methylmalonyl-specific AT of the Er PKS with that of the ethylmalonyl-specific AT of the Nd PKS. This substitution alone, however, was not sufficient to produce an Er derivative with an ethyl side chain; additional manipulations of the carboxylic acid precursor pools were necessary to generate the desired compound.

MATERIALS AND METHODS

Strains, Plasmids, and Media. The wild-type Er producer is *S. erythraea* ER720 (8). Plasmid pWHM3 (9) is an *Escherichia coli–Streptomyces* shuttle vector that is maintained in *S. erythraea* only when it contains heterologous DNA for chromosomal integration. Plasmid pDPE81 is a derivative of pKAS37 (10) in which a 1.7-kb *Bgl*II fragment containing the hygromycin resistance marker was inserted into the *Bgl*II site of pKAS37. *E. coli* DH5α (Life Technologies, Gaithersburg, MD) was the host used for plasmid construction and isolation. Liquid cultures of *S. erythraea* strains were grown in SGGP (11) for production of protoplasts and SCM (5) for metabolite or enzyme analysis. Plate cultures of *S. erythraea* strains were grown on R3M medium (5). Thiostrepton (Ts)-resistant strains of *S. erythraea* were grown in 25 µg/ml and 10 µg/ml Ts for plate and liquid cultures, respectively.

Plasmid pEAT4 (Fig. 2A) was constructed as follows. Cosmid pAIBX85, a pWHM3 derivative containing DNA from modules 3 and 4 of the Er PKS (corresponding to nucleotides 979–9349; GenBank accession no. M63677), was used to clone DNA flanking eryAT4. The 5' flanking region was isolated by digesting pAIBX85 with *Msc*I and *Bst*EII (nucleotides 4247–6033), treating with the Klenow fragment of DNA polymerase, and ligating the fragment into the *Sma*I site of pUC19 to generate pUC/5' flank. An *Avr*II site was engineered 13 bp downstream of the *Bst*EII site by PCR amplification of a 306-bp region of DNA from the *Pml*I site (nucleotide 5739) to 12 bp 3' of the *Bst*EII site (nucleotide 6045). The engineered *Avr*II site does not change the Pro-Arg residues encoded by this region (Fig. 2B). A *Bam*HI site was also included on the PCR primer just downstream of the *Avr*II site. The resulting fragment was digested with *Pml*I and *Bam*HI and cloned into the *Pml*I/*Bam*HI site of pUC/5' flank, replacing the native se-
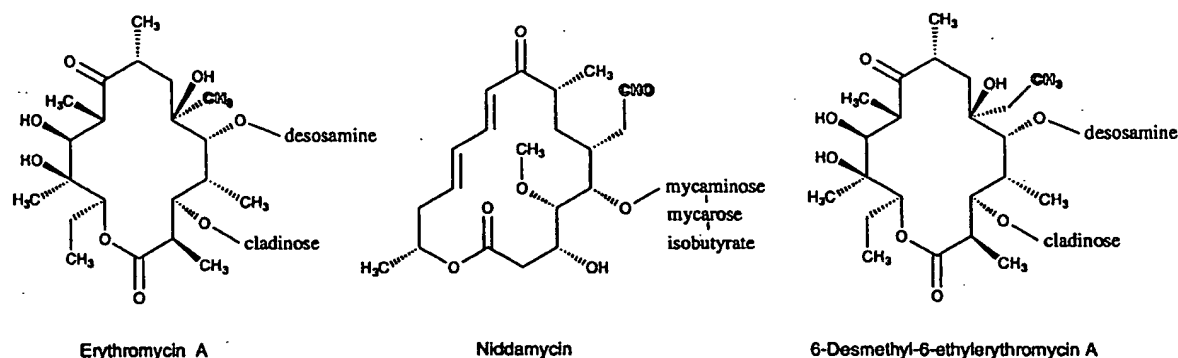
FIG. 1.   Structures of Er, Nd, and 6-ethylErA. Shadowed letters represent relevant side groups.

quence and resulting in vector pUC/5' flank/*Avr*II. The 3' flanking DNA was isolated by digesting pAIXB85 with *Pml*I and *Msc*I (nucleotides 6999 and 8977), treating with the Klenow enzyme, and ligating the fragment into the *Sma*I site of pUC19, generating .vector pUC/3' flank. nidAT5 was isolated by PCR amplification of cosmid 13f5 (7), which contains DNA spanning module 5 of the Nd PKS cluster. The 5' oligonucleotide was designed to create an *Avr*II site 12 nucleotides upstream of the beginning of the nidAT5 domain (nucleotide 25290; GenBank accession no. AF016585) and to place the amino acid sequence Pro-Arg-Lys-Pro in front of nidAT5 to correspond to the sequence that is found upstream of the eryAT4 domain. The 3' oligonucleotide generated an *Fse*I site at the end of the nidAT5 domain (nucleotide 26284, GenBank accession no. AF016585), resulting in a conservative Val to Ala change. A *Bam*HI site was also incorporated after the *Fse*I site. The fragment was digested with *Avr*II and *Bam*HI and ligated into *Avr*II/*Bam*HI-digested pUC/5' flank/*Avr*II, creating vector pUC/5' flank/nidAT5. The 3' flanking DNA was then cloned onto the 3' end of nidAT5 by first digesting pUC/3' flank with *Fse*I and *Bam*HI, gel purifying the 1920-bp

fragment, and ligating it into *Fse*I/*Bam*HI-digested pUC/5' flank/nidAT5, creating vector pUC/nidAT5/C6-flank. To generate pEAT4, the nidAT5/flanking DNA cassette was isolated from pUC/nidAT5/C6-flank by digestion with *Eco*RI and *Hin*dIII and then ligated to *Eco*RI/*Hin*dIII-digested pWHM3.

Plasmid pDPE-ccr (Fig. 3) was constructed as follows. The *Streptomyces collinus* crotonyl-CoA reductase (Ccr) gene (*ccr*) was subcloned from plasmid pZYB3 (12) by digestion with *Xba*I and *Bam*HI, which releases *ccr* and the upstream T7 ribosomal binding site. This fragment was treated with the Klenow enzyme and ligated into the polylinker of pDPE81 that had been digested with *Eco*RI and treated with the Klenow enzyme. The polylinker of pDPE81 is in the center of a 10-kb fragment of *S. erythraea* chromosomal DNA, which directs integration into the chromosome at a site that does not seem to affect Er production. Plasmid pDPE-ccr was designed so that the *ccr* gene expressed from the *ermE** promoter (13) and the Ts-resistance marker are left behind in the chromosome following a double-crossover event.

Genetic Manipulations. Standard molecular biology techniques were performed as described (14). Enzymes and re-. agents were purchased from Life Technologies. Protoplast transformation and marker replacement in *S. erythraea* were performed as described (5). One microgram of plasmid DNA was routinely used for protoplast transformations. Putative transformants were grown in SGGP medium containing 10 µg/ml Ts to confirm resistance. For chromosomal eviction of pEAT4, transformants were passaged twice in SGGP medium without Ts, cells were plated for spores, and individual colonies arising from spores were screened for Ts sensitivity.
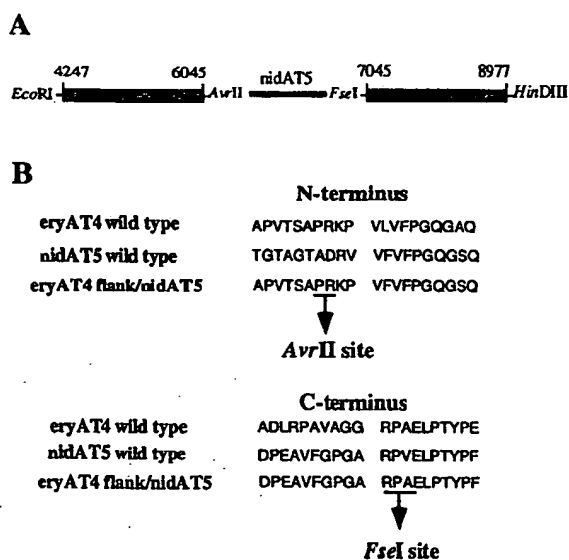
A

B

	N-terminus	
eryAT4 wild type	APVTSAPRKP	VLVFPGQGAQ
nidAT5 wild type	TGTAGTADRV	VFVFPGQGSQ
eryAT4 flank/nidAT5	APVTSAPRKP	VFVFPGQGSQ

*Avr*II site

	C-terminus	
eryAT4 wild type	ADLRPAVAGG	RPAELPTYPE
nidAT5 wild type	DPEAVFGPGA	RPVELPTYPF
eryAT4 flank/nidAT5	DPEAVFGPGA	RPAELPTYPF

*Fse*I site

FIG. 2.   (*A*) Diagrammatic representation of insert in vector pEAT4. DNA encoding the AT from module 5 of the Nd PKS cluster (nidAT5) was cloned between DNA flanking the 5' and 3' boundaries of the AT coding region in module 4 of the Er PKS cluster. The numbered boxes correspond to *eryA* sequence coordinates from GenBank accession no. M63676. (*B*) Amino acid comparisons in the junction regions of eryAT4, nidAT5, and the eryAT4 flank/nidAT5 construct. The amino acids encoding the *Avr*II and *Fse*I sites are indicated.
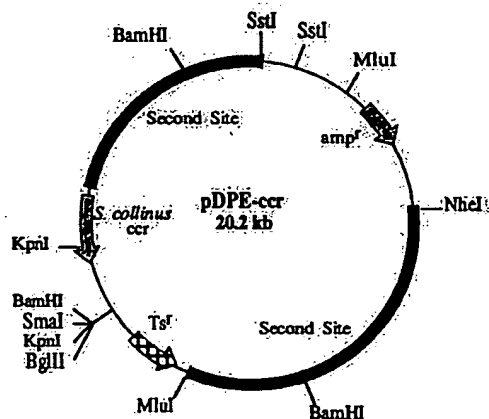

FIG. 3.   Plasmid pDPE-ccr. The thick black lines represent *S. erythraea* DNA, which allows for integration of the plasmid by homologous recombination into the *S. erythraea* chromosome. Also indicated are genes for ampicillin (amp) and Ts resistance.

Electroporation was performed as described (15) by using 1 μg of pDPE-ccr DNA. Because the plasmid is unstable when integrated into the *S. erythraea* chromosome, two consecutive platings on Ts-containing R3M plates results in resolution and eviction of plasmid sequences. PCR and Southern hybridization were as described (7).

Ccr Assay. Seed cultures of *S. erythraea* grown in SCM and *S. collinus* grown in medium A (16) were diluted 5-fold into SCM and medium A, respectively, and grown for 48 h at 30°C. Cells were harvested, opened with a French pressure cell, centrifuged to obtain cell-free extracts, and assayed spectrophotometrically for Ccr activity as described (16). One unit of Ccr activity is defined as 1 μmol of NADPH oxidized in 1 min.

Isolation and Identification of Metabolites. Small-scale (milliliters) isolation of metabolites from *S. erythraea*, TLC, and MS analysis were performed as described (5). Large-scale isolation for compound identification was as follows. Fermentations were conducted in 42-liter LH fermentation series 2000 stainless steel vessels (LH Fermentation, Maidenhead, UK). Each fermentor was charged with 30 liters of a medium consisting of 1.5% soluble starch, 2.2% soybean flour, 0.2% CaCO$_3$, 0.15% brewer's yeast (Wind Gap Farms, Baconton, GA), 0.102% MgSO$_4$·7H$_2$O, 0.0027% FeSO$_4$·7H$_2$O, and 1% soybean oil. Sterilization was at 121°C and 15 psi (1 psi = 6.89 kPa). A two-step seed protocol was used. Vegetative growth from a frozen stock of *S. erythraea* strain EAT4-ccr was inoculated at 1.5% into a 500-ml Erlenmeyer flask containing 100 ml of the following medium: 1.5% glucose monohydrate/1% soluble starch/1% Soytone (Difco)/0.9% yeast extract (Difco)/0.21% Mops/0.001% Ts. For the second step, a 2-liter Erlenmeyer flask containing 600 ml of the same medium was inoculated at 5% with the first-step growth. Both steps were incubated at 28°C on a rotary shaker at 225 rpm for 48 h. The fermentor was inoculated at 5% with the second-step seed growth. Fermentor temperature was controlled at 32°C, agitation was 250 rpm, aeration was 1 vol·vol^{-1}·min^{-1}, and head pressure was 6 psi. Silicone antifoam was added at 0.01% initially and was available on demand. Harvest was at 108 h.

The fermentation broth (28 liters) was filtered, the pH was adjusted to 9 with NH$_4$OH, and the broth was extracted with CH$_2$Cl$_2$ (twice with 14 liters each). The combined extracts were concentrated, and the residue was partitioned between the two phases of the heptane/methanol/0.02 M K$_2$HPO$_4$ system (pH 6; 1:1:1). The aqueous phase was evaporated, and the residue was dissolved in 0.05 M aqueous potassium phosphate buffer (100 ml), adjusted to pH 9 with NH$_4$OH, and extracted twice with equal volumes of CH$_2$Cl$_2$. The CH$_2$Cl$_2$-soluble fraction (1.65 g) was separated by droplet countercurrent chromatography (100 vertical columns; 0.4 × 24 cm) by using the upper layer from the hexane/ethyl acetate/0.02 M K$_2$HPO$_4$ (pH 8; 1:1:1) system as the mobile phase. The fractions were analyzed by bioassay against *Staphylococcus aureus* and by ^{1}H NMR. The bioactive fractions that were shown to contain macrolides by NMR were pooled, dried (0.4 g), and chromatographed on a Sanki Engineering (Kyoto) HPLC centrifugal partition chromatograph by using the hexane/ethyl acetate/0.02 M K$_2$HPO$_4$ (pH 6; 1:1:1) system. The active fraction was further purified by HPLC on C$_{18}$-derivatized silica by using an acetonitrile/methanol/0.01 M (CH$_3$)$_4$NOH/0.05 M KH$_2$PO$_4$ (pH 6; 73:10:59:59) system. Two-dimensional NMR experiments, including double-quantum correlation, heteronuclear multiple quantum correlation, and heteronuclear multiple-bond correlation, performed in C^{2}HCl$_3$, were used for structure elucidation and the complete assignment of the ^{1}H and ^{13}C NMR signals.

In Vitro Antibacterial Activity. Antibacterial activity was determined by the broth microdilution method (National Committee for Clinical Laboratory Standards (Villanova, PA) M7-A4, 1997), except that Brain Heart Infusion (Difco) broth was used as the test medium. Assays were incubated overnight at 35°C. Minimal inhibitory concentrations were defined as the lowest drug concentration (μg/ml) inhibiting bacterial growth.

RESULTS

Construction of *S. erythraea* EAT4. Plasmid pEAT4 was constructed to replace DNA encoding the AT domain of module 4 in the Er PKS (eryAT4) with DNA encoding the AT domain from module 5 of the Nd PKS (nidAT5) (7). *S. erythraea* ER720 protoplasts were transformed with pEAT4 DNA, and 10 transformants were obtained. Genomic DNA was extracted from one of the transformants for Southern analysis, in which probing with pWHM3 DNA confirmed the integration of pEAT4 at the appropriate location in the chromosome (data not shown). Nonselective growth to allow plasmid eviction through a double-crossover event yielded 96 colonies, of which 9 were found to be Ts sensitive. Southern analysis of the 9 clones showed that 3 had nidAT5 DNA sequences in place of eryAT4 chromosomal sequences, and the remaining 6 had segregated to wild type (data not shown).

Characterization of *S. erythraea* EAT4. To analyze the Er derivatives produced by *S. erythraea* EAT4, the three isolates and four of the wild-type segregants were grown in SCM medium for 4 d. The four wild-type segregants produced spots indistinguishable in color and R_f from the ErA standard. The three strains in which nidAT5 replaced the AT of module 4 of the Er PKS produced spots similar in R_f and color to the wild-type segregants, but the spots were much less intense (data not shown).

To determine the mass of the compound produced by *S. erythraea* EAT4, the supernatant of a 50-ml SCM culture of one of the isolates was extracted with ethyl acetate. The extract was subjected to TLC, but only the edges of the plate were sprayed with anisaldehyde to locate the region of interest. A 1-cm band of resin was scraped from the unsprayed portion of the plate at the R_f of ErA. The resin was extracted twice with 500 μl of an ethyl acetate/methanol (2:1) solution, and the organic phase was dried and then analyzed by electrospray ionization MS. Surprisingly, the mass of the protonated molecular ion of the compound was observed at *m/z* 734, which corresponds to the mass of ErA, not an ethyl-substituted ErA derivative.

Feeding of Ethylmalonyl-CoA Precursors to *S. erythraea* EAT4. One hypothesis for the failure of *S. erythraea* EAT4 to produce an ethyl-substituted derivative is that ethylmalonyl-CoA, the substrate for the nidAT5, is lacking. To test this hypothesis, cells were grown for 4 d in SCM medium containing precursor compounds, and ethyl acetate extracts of the cultures were analyzed. When either 50 mM butanol or 50 mM butyrate was added to the medium, a second spot running slightly faster than the R_f of ErA was seen. Butyrate-fed cultures (Fig. 4A) produced about a 1:1 ratio of the two spots, whereas the butanol-fed culture produced more of the ErA-like spot (data not shown). Addition of 10 mM ethylmalonate failed to produce significant amounts of either of the two spots described above. However, addition of 10 mM diethylethylmalonate was found to yield much more of the faster migrating compound than that seen with the other precursor compounds tested and very little compound migrating at the R_f of ErA (Fig. 4A). Cells grown in unsupplemented SCM medium produced only material migrating like ErA.

To characterize the newly synthesized compound, extracts of the butyrate- and the diethylethylmalonate-fed cultures were subjected to TLC, the region of interest was scraped from the plate and reextracted for electrospray ionization MS. The results showed that the butyrate-fed culture of *S. erythraea* EAT4 produced approximately equal amounts of compounds with protonated molecular ions at *m/z* 734 and 748. The 748 species is consistent with an additional methylene group on ErA, e.g., with an ethyl group replacing a methyl group on the
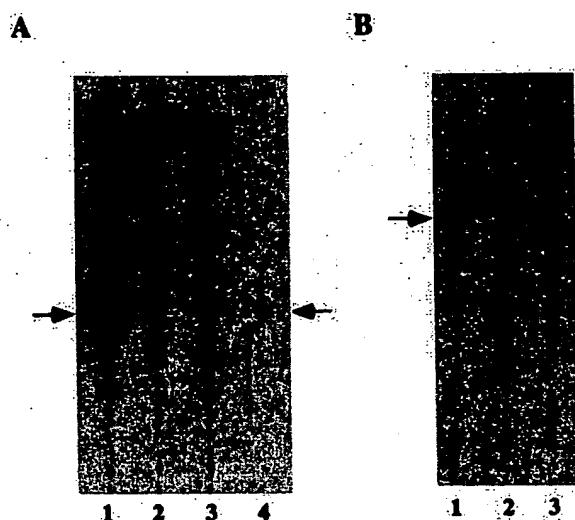
A   **B**

FIG. 4.   TLC analysis of S. erythraea EAT4 and S. erythraea EAT4-ccr. (A) Ethyl acetate extracts of S. erythraea EAT4 cultures grown under various conditions. Lanes: 1, SCM medium; 2, SCM + 50 mM butyrate; 3, SCM + 10 mM diethylethylmalonate; 4, ErA standard (5 μg). The arrows indicate the position of ErA. (B) Ethyl acetate extracts of S. erythraea strains EAT4 (lane 2) and EAT4-ccr (lane 3) grown in SCM medium, along with 5.0 μg of ErA standard (lane 1). The arrow indicates the position of ErA.

macrolide ring. The diethylethylmalonate sample contained primarily a compound with a mass of 748, with only a trace of ErA (734) present.

Genetic Manipulation of Ethylmalonyl-CoA Levels in S. erythraea EAT4. Successful production of an Er derivative with a mass consistent with the addition of an ethyl side chain by butyrate- and diethylmalonate-fed S. erythraea EAT4 suggested that the levels of ethylmalonyl-CoA, the likely substrate of the nidAT5, must be limiting. In streptomycetes, Ccr catalyzes the last step in the reductive biosynthesis of butyryl-CoA from two molecules of acetyl-CoA (12). Because butyryl-CoA can then be carboxylated to form ethylmalonyl-CoA, the possibility that Ccr could be used to increase the levels of the ethyl-substituted Er derivative was investigated. The ccr gene of S. collinus was expressed from the strong ermE* promoter at a site unlinked to the Er biosynthetic cluster in the S. erythraea EAT4 chromosome. Plasmid pDPE-ccr was electroporated into S. erythraea EAT4. About 40 Ts-resistant colonies were obtained, of which 7 were confirmed to be stable transformants. Genomic DNA was isolated from 2 of the transformants for Southern analysis. Probing with the S. collinus ccr gene confirmed that both isolates carried the gene at the expected location in the chromosome (data not shown). The strain was named S. erythraea EAT4-ccr.

To assess expression levels of the ccr gene, Ccr activity of S. erythraea EAT4-ccr was compared with that of S. erythraea EAT4 (parental strain) and wild-type S. collinus. We found (Table 1) that, although there was no detectable activity in the S. erythraea strain without the ccr gene, S. erythraea EAT4-ccr

had about 20 times the relative activity of the wild-type S. collinus strain from which the gene originated.

To determine the effect of the ccr gene on production of an ethyl-substituted Er derivative, ethyl acetate extracts of supernatants of 4-d SCM-grown cultures of the EAT4 strains with and without the ccr gene were examined (Fig. 4B). TLC analysis of 10 ml of extracted cells demonstrated that EAT4 without the ccr gene produced a spot that migrated with ErA, whereas EAT4-ccr produced a spot that migrated faster than the ErA spot, with no ErA-like material visible.

Extracts were also analyzed by electrospray ionization MS. The EAT4 strain produced a compound with a protonated molecular ion at m/z 734, indicating ErA production. The EAT4-ccr strain produced a compound with a mass of 748, which is consistent with the production of the 6-ethylErA. No compound with a mass of 734 (ErA) was detected in extracts of this strain.

Structural Analysis and Biological Activity of Compound Produced by S. erythraea EAT4-ccr. The structure of 6-ethyl-ErA was confirmed by spectroscopic methods. The high resolution fast atom bombardment MS analysis of the sample gave an M + H$^+$ ion at m/z 748.4846 [calculated for $C_{38}H_{70}NO_{13}$ m/z = 748.4842 (Δ 0.5 ppm)], indicating that this molecule has an additional methylene function compared with ErA. The presence of an ethyl moiety in the molecule at the C-6 position and its relative stereochemistry were confirmed

Table 1. Ccr activity in S. collinus and engineered S. erythraea strains

Sample	Enzyme activity, milliunits/ml	Protein concentration, mg/ml	Specific activity, milliunits/mg
S. erythraea EAT4	<0.5 ± 0.0	3.0	<0.2 ± 0.0
S. erythraea EAT4-ccr	46.9 ± 2.0	3.0	15.6 ± 0.7
S. collinus wild type	5.5 ± 1.0	7.0	0.8 ± 0.1

Table 2.   NMR data of 6-ethylErA

Carbon no.	^{13}C shift (δ), ppm	^{1}H shift (δ), ppm
1	175.5	
2	44.8	2.93
3	80.7	3.79
4	37.6	1.93
5	81.6	3.86
6	76.7	
7	38.4	1.72, 193
8	44.7	2.70
9	222.4	
10	37.7	3.01
11	68.9	3.82
12	74.4	
13	76.7	5.06
14	21.2	1.47, 1.91
15	10.7	0.84
2-CH₃	15.9	1.29
4-CH₃	9.7	1.14
6-CH₂-CH₃	27.6	1.70, 2.02
6-CH₂-CH₃	7.7	0.98
8-CH₃	18.6	1.17
10-CH₃	12.3	3.01
12-CH₃	68.9	3.82
1'	102.5	4.48
2'	72.5	3.21
3'	65.3	2.44
4'	28.3	1.22, 1.63
5'	68.7	3.52
6'	21.6	1.22
3'-N(CH₃)₂	40.7	2.29
1″	96.2	4.92
2″	35.3	2.34, 1.66
3″	72.4	
4″	77.7	3.02
5″	66.3	3.98
6″	21.6	1.29
3″-OCH₃	49.7	3.33
3″-CH₃	21.4	1.26

Numbering of carbon atoms of Er is as described (17).

by two-dimensional NMR. The stereospecificity of the ethyl moiety was found to be the same as that of the methyl group present at C-6 of ErA (Fig. 1). The ^{1}H and ^{13}C NMR assignments are shown in Table 2.

Antibacterial activity of 6-ethylErA was compared with that of ErA against a panel of *S. aureus, Staphylococcus epidermidis, Streptococcus pyogenes,* and *Enterococcus faecium* strains from the Abbott culture collection. Although 6-ethylErA does retain biological activity, it was found to be 15 to 60 times less potent than ErA against these strains. Representative minimal inhibitory concentrations (μg/ml) for 6-ethylErA and ErA, respectively, were as follows: *S. aureus* (4 and 0.06), *S. epidermidis* (4 and 0.12), *E. faecium* (4 and 0.12), and *S. pyogenes* (1 and 0.06).

DISCUSSION

In this work we have shown that an ethylmalonate AT can be substituted for a methylmalonate AT to produce a hybrid PKS that synthesizes the predicted macrolide product, 6-ethylErA. This suggests that Er analogs containing ethyl substitutions at positions other than C-6 or that ethyl analogs of other macrolides, such as rifamycin or rapamycin, can be prepared in a similar fashion by using the ethylmalonate-specific AT from the Nd PKS.

One key feature for the successful generation of 6-ethylErA was availability of ethylmalonyl-CoA as substrate for the ethylmalonate-specific AT of the Nd PKS in *S. erythraea.* This AT clearly demonstrates a relaxed specificity by reacting with both ethylmalonyl-CoA and methylmalonyl-CoA. In *S. erythraea* EAT4, the desired product could be obtained only by boosting the levels of ethylmalonyl-CoA by either precursor feeding or expression of Ccr, an enzyme involved in butyryl-CoA biosynthesis. These results suggest that ethylmalonyl-CoA derived from butyryl-CoA is not available at sufficient levels to compete with methylmalonyl-CoA for the ethylmalonate-specific AT. Ccr activity is not detectable in *S. erythraea* and attempts to detect a *ccr* gene in this organism by hybridization with the *S. collinus ccr* gene have failed (results not shown). In contrast, the *ccr* gene appears to be present in most streptomycetes (16), and in certain cases it appears to be clustered with antibiotic biosynthetic gene clusters. For example, a *ccr* homolog has been located in the biosynthetic cluster for tylosin (18), another 16-member macrolide with an ethyl side chain. Southern analysis of cosmid clones containing Nd biosynthetic genes also indicates that a *ccr* homolog may reside close to the Nd PKS (unpublished results).

It was demonstrated previously (5) that production of desmethyl Er derivatives by replacement of methylmalonate-specific ATs of the Er PKS with malonate-specific ATs from *Streptomyces hygroscopicus* and *Streptomyces venezuelae* was successful only in Er modules 1 and 2 (corresponding to C-12 and C-10 of Er, respectively) and did not give rise to a detectable polyketide when placed in module 4. In contrast, we successfully produced ethyl-substituted Er by replacement of the methylmalonate-specific AT of Er module 4 with the ethylmalonate-specific AT from the Nd cluster. In fact, in *S. erythraea* expressing the *S. collinus ccr* gene, replacements in modules 1 through 4 with the ethylmalonate-specific AT produced compounds with electrospray ionization mass spectra consistent with production of ethyl substitutions for methyl groups at C-12, C-10, C-8, and C-6, respectively. However, those in modules 5 and 6 did not produce compounds of mass 748 (data not shown). It is not known whether failure to produce the predicted compounds was caused by physical distortion of the PKS or its mRNA, inability of the PKS to process certain altered growing chains, or some structural

instability of the macrolide itself because of the introduced change. Thus, even though genetic information may be present for the production of novel compounds, it is still not possible to predict which substitutions will yield detectable levels of product.

It has also been demonstrated that the malonyl-specific ATs can be distinguished from methylmalonyl-specific ATs through sequence alignments (3). By using this strategy, Ruan *et al.* (5) predicted that an AT from an unidentified PKS of *S. hygroscopicus* was a malonyl-specific AT, and they used it successfully to produce desmethyl Er derivatives. The Nd ethyl-specific AT was found to cluster with methylmalonyl-specific ATs (7) in similar AT alignments. This similarity may explain why methylmalonate was used as a substrate by nidAT5 in *S. erythraea* EAT4 to produce ErA when ethylmalonyl-CoA was not available.

Finally, structural determination by NMR not only confirmed the presence of the ethyl side chain on the Er derivative produced by *S. erythraea* EAT4 but also showed that the absolute configuration at this chiral center is the same as that in ErA. In Er biosynthesis, epimerization is required at C-6 because the PKS uses (2*S*)-methylmalonyl-CoA for chain extension (19). It is unclear whether the cognate epimerization occurs when ethylmalonate is used in the synthesis of 6-ethylErA because the stereochemistry of the ethylmalonate incorporated at C-6 is not known.

1. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. (1990) *Nature (London)* 348, 176–178.
2. Donadio, S. & Katz, L. (1992) *Gene* 111, 51–60.
3. Oliynyk, M., Brown, M. J. B., Cortes, J., Staunton, J. & Leadlay, P. F. (1996) *Chem. Biol.* 3, 833–839.
4. Kuhstoss, S., Huber, M., Turner, J. R., Paschal, J. W. & Rao, R. N. (1996) *Gene* 183, 231–236.
5. Ruan, X., Pereda, A., Stassi, D., Zeidner, D., Summers, R., Jackson, M., Shivakumar, A, Kakavas, S. & Katz, L. (1997) *J. Bacteriol.* 179, 6416–6425.
6. Marsden, A. F. A., Wilkinson, B., Cortes, J., Dunster, N. J., Staunton, J. & Leadlay, P. F. (1998) *Science* 279, 199–202.
7. Kakavas, S. J., Katz, L. & Stassi, D. (1997) *J. Bacteriol.* 179, 7515–7522.
8. DeWitt, J. P. (1985) *J. Bacteriol.* 164, 969–971.
9. Vara J. A., Lewandowska-Skarbek, M., Wang, Y.-G., Donadio, S. & Hutchinson, C. R. (1989) *J. Bacteriol.* 171, 5872–5881.
10. Maine, G., Post, D., Satter, M. & Stassi, D. (1997) International patent application WO 97/06266 (2/20/97).
11. Yamamoto, H., Maurer, K. H. & Hutchinson, C. R. (1986) *J. Antibiot.* 39, 1304–1313.
12. Wallace, K. K., Han, L., Lobo, S. & Renolds, K. K. (1997) *J. Bacteriol.* 179, 3884–3891.
13. Bibb, M. J., White, J., Ward, J. M. & Janssen, G. R. (1994) *Mol. Microbiol.* 14, 533–545.
14. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
15. Fitzgerald, N. B., English, R. S., Lampel, J. S. & Vanden Boom, T. (1998) *Appl. Environ. Microbiol.,* 64, 1580–1583.
16. Wallace, K. K., Bao, Z.-Y., Dai, H. Digate, R., Schuler, G., Speedie, M. & Reynolds, K. (1995) *Eur. J. Biochem.* 233, 954–962.
17. Larty, P. A. & Faghih, R. (1993) in *Recent Progress in the Chemical Synthesis of Antibiotics,* eds. Lukacs, G & Ohno, M. (Springer, New York), p. 121.
18. Gandecha, A. R., Large, S. L. & Cundliffe, E. (1997) *Gene* 184, 197–203.
19. Weissman, K. J., Timoney, M., Bycroft, M., Grice, P., Hanefeld, U., Staunton, J. & Leadlay, P. F. (1997) *Biochemistry* 36, 13849–13855.